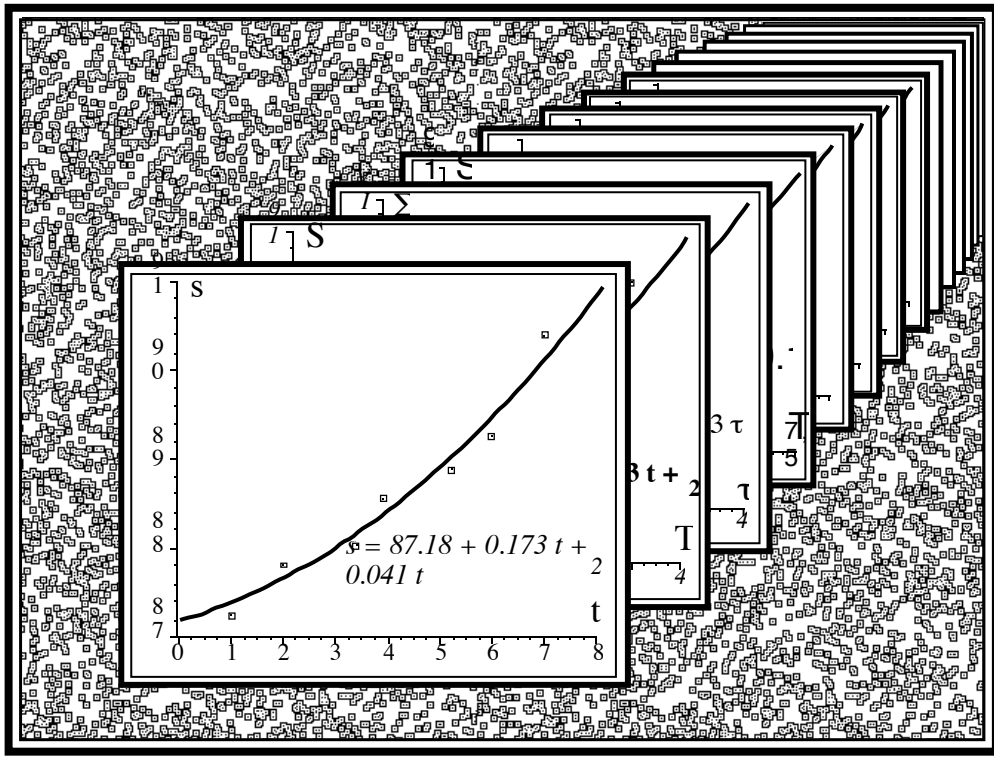
Representation and Change

A Metarepresentational Framework for the Foundations of Physical and Cognitive Science

by Francis Heylighen



About the book

This transdisciplinary work proposes a cross-fertilization between cognitive science and theoretical physics, within a framework inspired by systems theory. Cognitive science concepts are applied in an epistemological analysis of physical theories, considered as representations of change.

This analysis leads to the basic concept of *distinction conservation*, which appears necessary and sufficient to demarcate *classical* representations (classical mechanics) from *non-classical* ones (quantum mechanics, relativity theory and thermodynamics). It is observed that the most important cognitive and physical processes are non-classical (i.e. do not conserve distinctions), whereas the paradigms used for modelling and interpreting them are basically classical. This anomaly produces conceptual problems, exemplified by the paradoxes of quantum mechanics. The construction of an *adaptive metarepresentation* is proposed in order to solve these problems. This is a general framework for the representation of not distinction-conserving processes and representation changes. Finally a first sketch of a metarepresentational formalism is presented.

The book is addressed to a broad audience of researchers from different backgrounds. It is written in a style which avoids technicality, explaining difficult mathematical and physical concepts in the most simple way. It will be especially stimulating for philosophers and systems theorists interested in the integration of theories, for cognitive scientists involved in the application of ideas from physics, and for physicists wishing to understand the epistemological foundations of their models.

About the author

Dr. F. Heylighen has a degree in mathematical physics from the Free University of Brussels (VUB), and has afterward specialized in systems theory and cognition. He is presently doing research at the VUB on the modelling of complex evolution, with applications in the computer-supported solution of complex problems.

Copyright © 1987, 1999 by Francis Heylighen
Non-profit reproduction and distribution of this text for educational and research reasons only is permitted, providing this copyright statement is included.

Referencing this text:

Original printed version:
Heylighen F. (1990): *Representation and Change. A Metarepresentational Framework for the Foundations of Physical and Cognitive Science*, (Communication & Cognition, Ghent, Belgium).

Web edition(1999):
http:// pcp. vub. ac. be/ books/ Rep&Change. pdf

Table of Contents

PREFACE

The present work is the result of the personal exploration of a domain which at first seemed vague and unlimited, but which then gradually developed a clear and distinct shape. My first systematic research efforts were directed towards the conceptual problems of quantum physics. However, I soon became aware of the fact that the traditional conceptual framework of physics is too restricted to tackle problems which are essentially of an epistemological nature. On the other hand, the fact that these problems subsisted for over half a century without any clear sign of progress in the attempts to solve them, made clear that no better conceptual framework was available in classical epistemology either. This insight led me to direct my attention to the most diverse theories and disciplines which seemed to propose an alternative set of concepts able to disentangle these fundamental problems.

From this confusion of models and ideas there slowly emerged a coherent framework: the "cognitive paradigm". It became clear to me that all the different problems which had subsequently attracted my attention were basically problems about the nature of cognition: how can we know, i.e. get information about and construct a representation of, external phenomena? However, the discipline which studies these problems, "cognitive science", is very young (as far as I know, the first time the term "cognitive science" appeared was in 1975; see Bobrow and Collins, 1975; however the main concepts of the approach began to emerge some twenty years earlier). Therefore its theories and models are still confuse and incomplete, though their development is very rapid.

The most fundamental shortcoming seemed to be the relation between the cognitive concept of "representation" and the problem of change: how to represent change, and how to change representations? No existing cognitive formalism seems able to answer these questions. On the other hand the representation of change (i.e. dynamics) is basically the aim of physics. In this way I was led back to my original domain of interest. I then attempted to make a synthesis of the fundamental ideas of physical and cognitive science. The result is the concept of an "adaptive metarepresentation", which provides the guiding framework for the present work (see sect 3.5).

Although the problems touched in this work are often of a philosophical nature, I would not like this research to be classified as mere philosophy. We should remember that the basic ideas of most of the actual scientific theories (including physics) were once considered to be purely philosophical speculation. However, the development of new techniques can lead very rapidly to a transition from philosophy to science. The most recent example is the transition from epistemology and philosophy of mind to Artificial Intelligence, which is due to the introduction of the computer as an instrument for the simulation of mental phenomena. The criteria that I would use for distinguishing science from philosophy are "formalization" and "operationalization" (see chapter 3). How the present approach leads to a formalization and operationalization is explained in chapter 11.

"If this research does not belong to the field of philosophy, then to what field does it belong?", you might want to ask. I would answer that it cannot belong to one of the traditional scientific disciplines, since its purpose is just to unify or to integrate existing disciplines. In this sense it is really "transdisciplinary" (see sect. 3.4).

However if you would insist, and ask under which label it should be classified, then I would choose the label of "systemics" or "systems theory" (see von Bertalanffy, 1968; Mesarovic and Takahara, 1975), which is closely related to the more application-oriented field of "cybernetics". Both the objective (a domain-independent theory of how to represent evolving systems) and the conceptual tools (information, adaptation, system vs. environment, state vs. structure,…) of systems theory are very similar to those of the present research. However, I have tried to incorporate some more specific ideas from theoretical physics and cognitive science in order to enrich the conceptual framework of classical systems theory.

The overall ideal of transdisciplinarity has placed me for several problems during the elaboration of this work. First, I had to discuss different fields in which I am not an expert. This means that I shall probably be found guilty of some inaccuracies by the people who are experts themselves. I hope this will not detract them from reflecting on the general ideas I propose.

The second difficulty is to communicate the concepts of these various disciplines to an audience of people who cannot be experts in all the fields involved. Since I hoped my audience would encompass researchers from a very broad range of disciplines, I had to assume that my readers would have only a limited background knowledge about most of the topics discussed. Moreover many of the subjects I wanted to treat (e.g. the interpretation of quantum mechanics, the structure of relativistic space-time, the formalization of intelligence) are reputed to be very difficult to understand or to explain.

In orde to tackle this problem, I have tried to keep my formulation as didactic as possible. In particular, I have provided each new topic with a long introduction, where its relation with the previously explained concepts is defined. People who are well-acquainted with the topic might be tempted to skip the introduction. However, I think it would be better to read it, because the problem is generally introduced in a way different from the traditional one, with a particular emphasis on those transdisciplinary concepts which form the backbone of my approach.

A second technique to promote readability is to limit the use of mathematics. However, certain ideas are very difficult to express without using a formal language. I have tried to place the mathematical sections at the end of the different chapters, so that they could be skipped without too much harm by the readers whose mathematical background is limited. The readers who would like to see a more mathematical approach, on the other hand, are referred to the last two chapters, where I show how the basic ideas of my work can be formalized without too much effort.(A second, more down-to-earth reason to limit the use of mathematics was that my text processor had only limited capabilities for treating formulas; this explains why the notation is sometimes

unconventional: I had to replace certain conventional symbols by alphanumeric characters.)

Acknowledgments

The basic responsibility for this work is my own; there is no one to be blamed for the eventual errors it might contain but me. However, there are several people who have directly or indirectly contributed to its realization.

First of all I should thank Prof. Jean Reignier who stimulated my interest in the foundations of quantum mechanics, and who gave me the opportunity to carry out my research in his department of theoretical physics. It was here that I met dr. Dirk Aerts, who had a great influence on my early research, although we often disagreed about the strategy to follow. Now I can say that it was he who opened me the way to a cognitive approach of the foundations of physics (although he himself is somewhat reluctant to use the word “cognition”). Further I wish to thank the other members of the TENA department, in particular Torgny Sjödin and Jo Decuyper, for their sympathy and help.

My “initiation” in the field of cognitive science was stimulated mainly by people from two research centers: the “Communication & Cognition” group in Ghent, which was founded by Prof. Leo Apostel, and the Brussels AI-lab, led by Prof. Luc Steels. First of all I wish to thank Leo Apostel for his warm friendship, his encouragement, and his willingness to listen to and to comment all my ideas, however strange they at first might sound. Further I want to thank Profs. Paula Burghgraeve, Marc De Mey and Jean-Paul Van Bendegem, for the many discussions we had and for their collaboration on transdisciplinary projects.

From the Brussels AI-lab I would first of all remember Gina Fierlafijn, whose friendship provided (and provides) me with an unlimited source of inspiration, even though the subject of our discussions was seldom scientific. I further wish to express my gratitude to the other members of the AI-lab for the stimulating atmosphere they succeeded to create at the VUB. In particular I wish to acknowledge the help of my old friend Erik Van Releghem, of Peter Beyls and of Pattie Maes, who provided me with valuable information.

Lastly I want to thank some people from the psychology faculty, mainly Prof. Jean-Pierre De Waele, Luc Van Langenhove and Prof. Erik Rosseel, who together with the others helped me to create the “Transdisciplinary Research Group”, and who showed in this way that transdisciplinary cooperation is more than just a dream.

Brussels, January 1987. Francis Heylighen

Preface to the printed publication

This work was originally written as a Ph.D. thesis in physics. It was defended before the commission of the Faculty of Sciences of the (Flemish) Free University of Brussels (VUB) on March 26, 1987, and was accepted "summa cum laude". Apart from the layout and some typing mistakes nothing has been changed between the thesis version (January 1987) and the present book version. Since the thesis had been written with the intent to reach an audience which is broader than the mere defense commission, it did not seem necessary to do any major editing in order to prepare it for publication.

The addendum (chapter 12) was written in February 1987. It contains a reformulation of the main results of the work in a more formal manner, and shows how basic structural properties of different physical theories can be deduced from a few general axioms. I have also included an additional list of references (after the original bibliography), containing papers which were written after the first version of the work, and which contain some elaborations of ideas in the work, especially in relation to autonomous systems and to the dynamics of distinctions.

Brussels, December 1989 F. H.

Preface to the electronic publication

Since the printed version of this book, published by Communication & Cognition in Ghent, Belgium, is no longer widely available, and since several people who read it told me how much they had appreciated the work, I decided to make the text available electronically on the World-Wide Web. Again, this new version is minimally different from the previous one. The main corrections concerned formatting and some grammar and stylistics. Nothing has been changed to the content, except that the list of “additional references” has been augmented with a few papers written since the previous publication. Most of these papers are available through the same website as this text itself, via the URL: http://pespmc1.vub.ac.be/Papers/.

Brussels, June 1999 F. H.

PROLOGUE: The conceptualization of change

The philosophical conception of change as the very fabric of reality can be traced back to Heraclitus, who is considered one of the first of the classical tradition of Greek philosophers. For him, all things and the universe are in a continuous, ceaseless flux, nothing exists as a static entity, only change is real. In order to show that no two experiences can be identical, he compared this endless process to a river, and remarked that "you can never step in the same river twice". A similar awareness of the pervasiveness of change and impermanence of things was at the base of Eastern philosophy, as expressed by Buddhism and Taoism.

Western thought, however, developed in a different fashion. The alternative view was expressed by the generation of Greek philosophers following Heraclitus. Its most radical proponents were Parmenides and his disciple Zeno, who maintained that change is an illusion, that there is just one timeless "*Being*", in contrast to the "*Becoming*" of Heraclitus.

It is this view which guided the beginning of Western science, first in its discovery of the "timeless truths" of logic and geometry, then, in the application of those mathematical truths to the description of mechanical motion. In order to apply these static descriptions to motion, i.e. a form of change, a more general mathematical framework had to be constructed. This was mainly due to Newton, who introduced the concept of Absolute Time as a one dimensional geometrical space, which could be coupled to the three dimensional space of positions by using the concepts of velocity and acceleration.

In order to describe velocity and acceleration in a mathematical fashion, and in this way to overcome the paradoxes proposed by Zeno to demonstrate the impossibility of motion, Newton (and independently of him, Leibniz) invented the differential calculus. This was the origin of the mechanistic paradigm, which dominated science for centuries.

Although now the existence and representability of a certain form of change, mechanical motion, were acknowledged, the fundamental concepts of the world view still emphasized permanence: the elements of the world which underwent this motion were seen as atoms or particles, i.e. as permanent pieces of matter; the movement itself took place in Absolute, unchanging Space and Time, according to the predetermined, eternal Laws of Nature. In this deeper sense, change is still an illusion: every event that will happen is determined a priori and could be said to already exist in some transcendental realm beyond our limited field of experience (call it the Mind of God, or the Four-dimensional Space-Time Continuum, or whatever you like).

The past century has seen a gradual erosion of this image of the universe which is based on absoluteness and permanence: the emphasis of scientific enquiry has moved back from "Being" to "Becoming" (cf. Prigogine, 1979; Prigogine and Stengers, 1979). A more profound analysis of the fundamental concepts of the

mechanistic framework has revealed that their apparent permanence is only approximate.

Elementary particle theories have shown that the microscopic constituents of matter are not permanent at all: particles are continuously created and destroyed through mutual interactions (or even, in the case of virtual particles, through "fluctuations of the vacuum"). On the macroscopic level, recent advances in fields such as thermodynamics, and biology point out that larger systems (crystals, stars, organisms, …) can and do undergo drastic transformations, which alter their most basic structures. According to the theory of general relativity, even the geometrical structure of space is continuously changing, and this leads to cosmological models in which the Universe itself and its "Laws of Nature" are periodically reprocessed (Misner, Thorne, Wheeler, 1974).

This reintroduction of change into the conceptual models of the world was paralleled by an increase of the rate of change in culture and society as a whole. It is well-known that the rate of innovation in our present-day society is faster than it ever was, and this has a profound influence on the psychology of the people who live in it (Cf.. Toffler, 1970). It is no surprise then, that we can no longer find people such as Parmenides, who would deny the reality of change.

But does this mean that there is not anything left that is permanent at all? In orde to get convinced about that, it suffices to study any given phenomenon, and to analyze it deeply enough so that its underlying dynamics becomes apparent.

Take for example a billiard ball (which is a classical illustration of the mechanistic model) at rest. As you look at it, it appears completely static, unchanging. Yet we know from solid-state physics, that it is composed of coupled molecules which are continuously vibrating and exchanging energy under the form of waves. This hidden thermal motion may become apparent when the temperature reaches some threshold so that the ball melts, or when the stress on the ball is raised so that it breaks apart. In both cases, the relative motion of the molecules becomes so large that their coupling is undone.

The pervasiveness of these hidden, internal processes becomes even clearer if we consider a more complex system, such as a human being. Physiology tells us that everywhere in our body, even when it does not move, continuous chemical processes are under way. Molecules are transported between the cells, are destroyed and created through chemical reactions, are exchanged with the outside world through eating or breathing. On the biological level, cells die and reproduce, organs direct and control physiological processes. So, inside the human body everything is in a ceaseless change. Yet, we still recognize the individual John Smith, even though we may guess that approximately every molecule in his body has changed since the last time we saw him.

What about more abstract entities? Consider for instance the concept of "truth". If it is acknowledged that some assertion, e.g. 2 + 2 = 4, is true, does this not mean that it remains true once and for all? In modern epistemology, the view was advanced that truth is a relative notion, that it depends on the larger framework or conceptual system in which the assertion is formulated. For example, mathematical truth is established by a "proof", which is a sequence of

deductions, following certain "rules of deduction", from a set of axioms. Hence, the truth of an assertion will depend on the axioms and rules which were postulated at the beginning. If the framework, and thus the axioms or rules, are changed, then also the truth of a given assertion may change. That in practice conceptual frameworks do change, is clearly shown by the history of science.

These examples, and their underlying ideas, that I have sketched, point to a world view which is radically different from that of Parmenides or Newton. In this world view nothing is permanent, everything is changing and the closer you look at it, the more it appears to change.

This world view may be frightening to many people. It seems that everything is confused, unpredictable, uncontrollable, that nothing can be trusted anymore. That a great amount of change in the life of an individual can lead to feelings of inadequacy or anxiety, to stress and hence to all kinds of mental and physical illness, is well-established (Toffler, 1970). On the social level, this phenomenon can lead to a generalized pessimism, to a crisis of belief and values, to a mentality which is expressed by the slogan: "No future".

On the other hand, this injection of change in our society and its (scientific) world view is greeted with enthusiasm by certain people. This view of the universe as a complex, evolving network of processes and interactions, reminds them of ancient Eastern and Western philosophies and mystical traditions. They consider the new scientific world view as the basis for an "emerging science of wholeness and becoming" (Briggs and Peat, 1984; Ferguson, 1980). The growing awareness of the dynamic and interactive character of the universe is for them the first sign of a new evolutionary transition, which will take mankind to a higher level of consciousness and integration with the changing universe.

Although both the pessimistic and the optimistic viewpoints, as sketched here, are directly inspired by scientific and technological innovation, none of them has a real scientific underpinning. Both are based on vague intuitions, rather than on a clear conceptual framework. In order to build up a new science of becoming you need more than an intuitive awareness of holistic and dynamic principles.

This leads us to a first formulation of the problem with which this work is concerned:

if the evolution of science and society compels us to reject the classical conceptual framework based on permanence, then how are we to replace it?

Although this question may sound rather metaphysical, it is really very pragmatic. To do things, to solve problems efficiently, to tackle new situations in an adequate way, you need some kind of understanding of the world around you, some framework to guide you in your decisions and in your planning. The more your environment is changing, the more difficult it will be to find such a guidance on which to rely, but also the more you will need one.

The success of the classical, Newtonian framework was that it allowed to plan, to anticipate certain processes in a very accurate way. The most spectacular illustration of that is that you can use it to put a man on the moon. Yet we know that its basic principles: the permanence of matter, of geometry and of natural law, can no longer be sustained. But does this mean that we have to throw it away completely?
The problem is the same as with our friend John Smith: if we know that all the matter in his body has changed since the last time we saw him, then must we conclude that he is no longer John Smith? What is happening here, is that if you conceptualize some phenomenon,—e.g. the geometry of space, John Smith, or the mathematical truth that 2 + 2 equals 4—thenou necessarily idealize it, you consider only a certain aspect of it which remains more or less invariant while you are using it, even though the phenomenon as a whole is continuously changing.

It is this process of idealization, of the extraction of certain invariant features from a changing situation, which allows the construction of a guiding framework for solving different kinds of problems in an efficient way, that is to say without getting lost in chaos and confusion. The result of this process may be called a "*representation*": a model, an abstract structure which represents changing phenomena in a stable way. The aim of science is just to construct such representations, and to make them as efficient as possible.

What was missing in the Newtonian paradigm was the awareness that it was nothing more than a representation, that it did not express Absolute and Eternal Truths, but that it just provided one way to represent rather accurately a very specific kind of change: mechanical motion. If you want to represent more radical types of change, such as those which determine the evolution of the present-day society, then you must look for an alternative for the mechanistic framework.

We can now formulate the basic research problem in a better way:

how can we generalize the classical mechanistic representation so that it can be used to represent more profound types of change?

In order to approach this problem we shall have to formulate it more explicitly. This means that we shall have to clarify the concepts which are used in its formulation. The two fundamental concepts are "*change*" and "*representation*".

The concept of "change" was discussed in this chapter. As we have shown, change is always and everywhere. Therefore I will not try to define it, but consider it as a primitive term. As was argued, the absolute negation of change, "permanence", does not exist. However, to contrast different levels of change, I will use the concepts of "*stability*" and "*invariance*, which are defined in a relative way: one phenomenon will be said to be stable relative to another phenomenon (e.g. its environment), if it does not change as rapidly or as easily; alternatively, it will be said to be invariant relative to a set of transformations or

processes, if none of these particular transformations makes it change. What this means in practice, will become clear through its use.

This leaves us with the second basic concept to define: representation. This problem will form the subject of the following two chapters.

CHAPTER 1: The concept of representation in physical and cognitive science

1.1 Physical science

The birth of modern science is traditionally traced back to Galileo. What distinguishes the work of Galileo from that of his predecessors, is that he combined consequently two approaches: *empirical observation* and *formal (mathematical) description*. This allowed him to model physical phenomena in a general and controllable way: the principles he inferred could be applied and checked in an accurate way by other people in quite different settings.

His approach can be seen as the beginning of *physical* science, which may be defined as the (scientific) representation of universal types of change, that is to say of those external processes which are general enough so as to be independent of the specific internal organization of the systems that undergo them. For example, the principles that Galileo derived concerning the motion of falling objects, are the same, independently whether the falling object is a stone, an apple, or a man.

Now in what sense does this approach lead to a "representation"? First, there is the mathematical structure, formulated for instance as an equation relating certain parameters, which provides a clear, explicit and manageable description of the phenomenon. Second, there is the operation of observation, which tells us just how this abstract structure is related to the phenomenon it describes. For example, it could state that a certain parameter in the equation corresponds to the distance measured with a meter stick, between the point where an object began to fall and the point where it reached the ground. So, in general we can say that a representation is an abstract structure which is related through certain operations with external, physical phenomena.

What is interesting about such a structure, is that it can be used to anticipate certain observations. For example, if you have measured the distance between the top of a tower, and the ground level below, then you do not have to use a clock in order to know how long it will take for an object to reach the ground. By using Galileo's equation you can deduce the duration of the fall from the height of the tower. What the representation tells you is that if you would measure the duration, then you would get the same result as the one deduced from the known height by using the equation. So, what the equation represents is the dynamical relationship between certain observations of physical phenomena.

The concept of representation defined in this way can of course be applied to the whole of physical science from Galileo until now: all theories and models of physics are representations, are mathematical structures relating empirical observations. This characterization is still quite vague, and not very practical. In mathematics and physics itself, the concept of representation is often used in a more specific sense.

For example, a mathematical physicist may speak about the representation of a vector, of an operator or of a group (see e.g. Saletan and Cromer, 1971). Abstracting from the purely technical definition, this means that he has some abstract concept which can be used to describe a dynamical phenomenon, but that it first has to be given a detailed form so that its relation with the concrete phenomenon can be clearly specified.

For example, the state of a system can be described by an abstract vector, but if you want to construct a manageable model of this system, then you must first express or "represent" this vector as a column of numbers. These numbers are called the components or coordinates of the vector in a certain basis, and can be interpreted as the empirical values obtained by making certain measurements on the system. As long as those numbers are not specified, the *operational relation* between the algebraic structure formed by the vectors and the concrete system cannot be established.

The same remark can be made about representation of groups: a group is an algebraic structure that can be used to describe possible changes of the state of a system. Yet, to make this description operational, this structure has to be embedded in a coordinated space, so that the action of each element of the group can be specified by a matrix, which transforms one column of coordinates into another one.

In order to differentiate this more restricted sense of the word "representation" from the general concept, I would rather use the word "*indexation*" instead. Indeed, what happens here, is that the abstract concept (e.g. vector or group) is indexed: it is labeled by coupling it to a set of indexes or pointers (e.g. coordinates) so that it can be easily located and related to the concrete phenomenon it is supposed to represent. The importance of this indexing function will become clearer when we shall discuss the concept of representation in the context of computer science.

There is yet another way in which the word "representation" can be used in the physical sciences. This third signification is somewhere in between the first, very general sense, and the second, rather restricted sense. As such, it comes closer to the sense we are looking for. It is the sense intended when scientists speak about the "*dynamical or state-space representation of a system*" (Mesarovic and Takahara, 1975; Gille and Clique, 1975). This approach is often used for the study of engineering systems, such as electrical circuits, or mechanical devices, but it can be applied to all kinds of physical systems.

The basic idea is that a certain system is changing, partly autonomously, partly under the influence of its environment. The scientific observer now wants to make a model of this evolution, so that he can anticipate what will happen if the system is subjected to a certain environmental input (which may be controlled by him). He supposes that this evolution can be represented as a sequence of states, such that at each time t, the system is characterized by a state s(t). This parameterized sequence forms a trajectory in the space of all possible states.

The "*representation problem*" can now be formulated as follows:

what form should be given to the states so that this trajectory and its observable manifestations can be easily computed?

In order to explain how this is done it is necessary to give a definition of "behavior" and "state" of a system. In this approach a system's behavior is characterized by the way it reacts (*output*) to the actions exerted upon it (*input*) by the environment or by the observer. Output and input are in general processes, i.e. "phenomena-in-time". The relation between input and output which defines the system as a dynamic entity, is in general not a function, i.e. for a given input process several different output processes are possible. This means that if one wants to predict unambiguously how the system will react to a determined input, he needs some extra information. This extra information, which summarizes in a certain sense the past history of the system before the prediction is made, is called the *state*.

More precisely: if one knows

(1) the general dynamical laws which constrain the behavior of the system, and
(2) the external actions to which it is submitted (input),

then the "state" will be the set of characteristics whose knowledge must be added to the knowledge of (1) and (2) in order to determine the further behavior of the system (Gille and Clique, 1975). The "state" in this sense is a conceptual construct, which is itself not directly observable, but which allows us to specify the relation between the observable input and output processes.

"Representing" the system then means that you introduce an intermediate formal structure, the state space, together with a dynamical law, so that with its help the observable reaction of the system to a known input can be efficiently computed. In practice, the problem amounts to a choice of state variables which makes the expression of the dynamical law, and its connection with the input and output variables, as simple as possible. The dynamical law then takes the form of a set of equations relating input, output and state variables as a function of time. If the equation can be easily solved, and if for a given input function and initial state, the solution (state and output function) is unique, then the representation may be called adequate.

Summarizing we can say that as yet the concept of representation has no clear and generally accepted meaning in physical science, but is used in several more specialized senses, depending on the context. Nevertheless, these different significations are sufficiently similar so that we can propose a definition which integrates them all:

a *representation* is a formal structure relating observable phenomena;
as such, it provides a mechanism for deriving anticipated observations (output) from actual observations (input);

its efficiency depends on its mathematical form, which must be sufficiently detailed to embody all information needed for an unambiguous solution of the problem, but not so complex as to make the search for this solution needlessly difficult;
the dynamical character of a phenomenon, i.e. the way observations change in time, is expressed in this formal structure by using a trajectory in a space of states.

1.2 From physical science to cognitive science

Whereas physical science can be defined as the study of nature, cognitive science might be characterized as the study of mind. However, this characterization is rather vague, and the mind-nature dichotomy is no longer adequate to categorize the objects of scientific enquiry. For example, should a machine, designed by man, be considered as a natural or as a mental phenomenon? (One way to avoid this dilemma, is to classify it as "artificial", cf. Simon, 1979). It will be shown later how the concepts of representation and change can be used to transcend this dichotomy.

For the moment, it suffices to call a science "*cognitive*" if it attempts to explain mental phenomena, such as perception, memory, thought, language, and even emotion and motivation (although this is seldom acknowledged as being dependent on cognition). The word "cognition" derives from the Latin verb "gnoscere" which means "to get knowledge of", hence "to grasp mentally". There are two aspects in this definition: a static (or stable) one, knowledge; and a dynamic one, the process through which this knowledge is acquired. Hence, the proper domain of cognitive science could be specified as:

the relations between the stable organization of the mind (cognitive structures), and the changes to which it is subjected through its interaction with the changing environment (cognitive processes).

Once again, we encounter the fundamental tension between stability and change.

Historically cognitive and physical science have evolved in a quite different manner. Physical science managed to acquire a clear, elaborated and reliable framework very early. Although this conceptual framework, which first appeared with Newton, has undergone several changes since, it is still generally acknowledged that all models of physical phenomena are in principle reducible to a few fundamental concepts and laws, which are formulated by theoretical physics. Thus theoretical physics, with its basic concepts such as space, time, particles, dynamical laws, …, appears as a central paradigm for all physical sciences.

Until recently, nothing of this sort existed for the sciences of the mind. Mental phenomena were investigated by such various disciplines as philosophy, psychology, linguistics, logic, anthropology, sociology, pedagogy, neurology, …, which had all their proper concepts and methodologies. Only a few decades ago some new ideas began to appear which seemed capable to create models of

mental phenomena characterized by a clear theoretical structure with operational implications. The tools which made this development possible, were the mechanical computer, and the simultaneously evolved concepts from cybernetics and information theory. The vagueness and ambiguity which characterized the existing conceptualizations of mental phenomena, could now be replaced by exact formalizations, which could be tested through experiments and computer simulations.

The synthesis of these new concepts and methods, inspired by the physical sciences, with certain more traditional ideas of the sciences of the mind, led recently to the emergence of an integrating framework, which may be called the "*cognitive paradigm*" (De Mey, 1982), and which provides the foundations for the new interdisciplinary field of "*cognitive science*" (Bobrow and Collins, 1975).

Two of its basic concepts are "*information processing*" and "*representation*". Whereas the first one stresses the dynamical aspect, the fact that external stimuli undergo complex transformations before their eventual "meaning" is extracted by the cognitive system, the second one points to the fact that in order to be efficient this processing needs some stable structure, called a "representation", to guide it. It is this representation which embodies the knowledge used by the cognitive system to interpret stimuli, to make decisions and to solve problems.

We will now try to analyze what this cognitive "representation" concept means in practice, and how it is related to the "representation" concept that was defined in the context of physical science.

1.3 The philosophical theory of ideas as representations of external objects

Historically, the study of cognitive phenomena began with philosophical speculations on the nature of mind and knowledge. One of the fundamental questions was how Mind and Matter are related. At first, the debate seemed to be a struggle between those people who maintained that only mental phenomena, "ideas", were real (idealism), and those who stated that only material objects were real (materialism). As a way of synthesis (or rather of compromise), Descartes proposed a dualistic philosophy, in which both Mind and Matter were accorded independent existence.

The remaining problem was to specify how these two realms can interact. Clearly, our mind can sense, can think about, can act upon material entities. With respect to action, Descartes gave no clear solution to the problem, but with respect to sensation and thinking, he proposed that the mind does not know its objects directly but only through the mediation of ideas which represent them. This theory of "representative ideas" had followers among such different philosophers as Hobbes, Locke and Berkeley.

The basic intuition is that if we perceive or think about some object, then what is present in our mind is not the object itself, but an image or representation, which is a kind of mental or ideal ambassador of the real, material object. Naively, the process through which these representations come

into being can be visualized as follows: a material object which is perceived induces some kind of patterned excitation of the sense organs (e.g. the image formed at the back of the retina during visual perception); this "image" is apprehended immediately by the consciousness (direct perception), but at the same time it leaves a trace in memory (*engram*); if we think about an object which is not directly present to the senses, then we merely revive this engram, and so the image of the object reappears before our mind.

The problem with this simplified picture is that it leads to the idea that a cognitive system is not very different from a *photographic camera*: the patterns of light coming from the object (stimuli) pass through some system of lenses (sense organs), and are projected onto a screen (consciousness), were they leave a permanent trace in the photographic emulsion (memory).

A first difficulty is that this view implies that our ideas should in some sense resemble, be *isomorphic* to the phenomena they are supposed to represent. This is clearly not the case as far as more abstract ideas are concerned. E.g. if I think about the number 8673, I clearly do not see 8673 dots before my mind's eye; there is no structural similarity between the concept 8673 and some collection of 8673 objects I once might have seen before me.

But the isomorphism of phenomena and ideas is not even true for concrete visual imagery. This can be illustrated by an anecdote reported by the philosopher Alain:

> once a friend of Alain told him that if he closed his eyes, he could clearly imagine a view of the Parthenon. Alain answered that if this mental image were so vivid, then his friend should have no trouble counting how many columns were in front of the Parthenon. Since his friend obviously was unable to do that, Alain concluded that whatever mental representation of the Parthenon his friend seemed to experience, it clearly was very different from a photographic image.

A second, even more serious difficulty raised by the "photographic camera" view of the mind is that of the *homunculus*. If perception is nothing but the projection of images onto a screen, and memory not different from a set of photographic prints of those projections, then who is looking at the screen, and shuffling through the photographs? The only way to answer this question is by introducing a little man, a "homunculus", who is sitting somewhere in our brain, looking at different incoming and stored mental images in order to decide what they really mean, what should be done with them, to which aspects should be paid attention. This leads us to postulate a *mind within the mind*, and that, of course, does not bring us one step closer to solving the problem of how the mind really works.

These difficulties, together with other criticisms, led to a decline in popularity of the theory of ideas as representations of the outside world. This trend was reinforced by the appearance of behaviorist psychology, whose basic doctrine is that mental phenomena can only be studied indirectly, as associations between externally observable stimuli and responses; hence any speculation about the

structure of internal representations or ideas should be regarded as meaningless. This had as a consequence that by the middle of the twentieth century, the concept of representation had virtually disappeared from the existing theories of mind. Its reappearance was mainly due to the discovery of the computer as an instrument for the simulation of mental phenomena.

1.4 Artificial intelligence and knowledge representation

The fundamental belief characterizing the field of *Artificial Intelligence* (AI) is that the computer as a mechanical, information-processing device can be used to simulate processes which are usually associated with human intelligence, such as perception, understanding, learning, reasoning and problem solving. This belief is founded on the idea that all mental activities can be reduced to some form of *information processing*, and that, since information is a purely abstract quality, this process is independent of the detailed physical structure of the information processing system. Hence, electronic computer and organic brain alike can undergo basically the same kind of information processes.

In order to model an abstract entity such as information with the help of a concrete object such as a computer requires the use of symbols, i.e. phenomena which refer to something else than their purely physical substance. Here we already encounter the concept of representation in its most basic form: the symbol (e.g. a certain state of an electronic memory) represents a unit of information (e.g. the number 8673).

But to process complex information in an efficient way you need more than a collection of elementary representations (corresponding to information units), you need some global framework to guide the process at each decision point. It is this awareness which has gradually developed through the subsequent attempts to build intelligent computer systems, and which led to the concept of knowledge representation. This development was clearly summarized by De Mey (1982), who distinguishes four subsequent stages:

> "- a monadic stage during which information-units were handled separately and independently of each other, as if they were single, self-contained entities;
> a structural stage which considered information as a more complex entity consisting of several units arranged in some specific way;
> a contextual stage where, in addition to an analysis of the structural organization of the information-bearing units, supplementary information is required to disambiguate the meaning of the message;
> a cognitive, or epistemic, stage in which information is seen as supplementary or complementary to a conceptual system that represents the information processor's knowledge or model of his world. "

These stages can be illustrated by the problems encountered in the computer simulation of *perception* and *language understanding*. Both these approaches attempt to model how an intelligent system interprets, i.e. attaches meaning to stimuli (sensory or linguistic):

In the first, monadic stage, the stimulus is decomposed into *simple units* (e.g. spots of light, or separate words), to which some fixed meaning is attached.

In the second stage, it is recognized that meaning resides not only in the separate units, but also in the *structure or pattern* they form (e.g. a sentence formed according to syntactical rules, from separate words). But in practice this analysis still does not seem sufficient to attach meaning to certain more ambiguous stimuli.

Therefore, in the third stage, another factor is taken into account: the *context* in which the stimulus is received, that is to say that extra information is needed, which is not in stimulus itself, but which helps to situate it, to determine to what it is referring (e.g. to understand the meaning of the sentence "He is nice", you must know from the context to whom the pronoun "he" is referring). The problem with a contextual approach is that it is very difficult to define the context, to make explicit what is in the context and what is not.

This leads to the fourth stage, where each stimulus is situated within a *conceptual system or model* that specifies which meaningful signals can or cannot be expected in the given situation.

This system is called a "representation" of the information processor's knowledge of his environment or situation. The basic idea is that the meaning is not so much in the stimulus itself, but in the representation. The stimulus only *selects* or activates certain prearranged *categories*, whose meaning is determined by the way they are interconnected. Whether the information processor has correctly interpreted or understood the stimulus can then be tested by asking questions about the perceived situation: if the information provided by the stimulus has been processed in an adequate way, the system should be able to give correct answers.

There are several reasons why this fourth, cognitive stage of information processing is much more powerful than the previous stages. The main advantage is *reduction of complexity*: in a realistic situation an intelligent system such as a human being is bombarded with stimuli, with information units; since the information processing capacity of the system is limited, it is unable to process in depth all these signals; therefore it needs some stable framework to organize these stimuli in categories so that only the most important stimuli will be processed.

The stimuli to which the system effectively pays attention are very few with respect to the whole. In order to carry out this selection the system needs a certain knowledge of the world in relation to itself, which allows it to anticipate, to infer from a few input data what important phenomena are to be expected. Then it can concentrate its attention or processing activity upon those aspects of the expected phenomena where a choice has to be made, and neglect all other stimuli which have no direct bearing upon these expected decisions.

To take a classic example from cognitive science: suppose that you enter a restaurant with the purpose of having a dinner. Then your "knowledge representation" of restaurant situations tells you to look for an empty table, to sit down, and to expect the coming of a waiter. Normally you will not pay much

attention to other people who have no waiter's clothes, or to other stimuli such as the color of the chairs, or the brightness of the lights. When the waiter appears, your attention will shift to the menu card he offers to you, and your knowledge of the tastes of different dishes will come into play to help you reach a decision about what to order. In the meantime the other stimuli impinging upon your senses (sounds from the traffic, other people talking, the hardness of your chair, …) will barely be processed.

When, however, something unusual happens, i.e. something which does not fit into the set of expectations generated by your knowledge representation of the actual situation (e.g. someone cries "Fire!") your attention to these other stimuli is reactivated, and another representation of a situation (e.g. the representation of what to do in case of emergency) takes over control.

Now, how does this concept of "representation" as it is used in the context of AI and the cognitive paradigm relate to the general definition of representation we are trying to construct? In fact when someone speaks about a knowledge representation he uses the word "representation" in two different but simultaneous senses. These two senses can be illustrated by two quotations from Bobrow (1975), where he uses the two different senses on the same first page of his introduction:

> "Workers in cognitive science have worried about what people know, and how to represent such knowledge within a theory."
> "I propose here a framework where representations are viewed as the result of a selective mapping of aspects of the world. Suppose we take a "snapshot" of the world in a particular state at some instant in time. Call this state world-state-1. Through some mapping M, a representation (call it knowledge-state-1) is created which corresponds to world-state-1. This corresponds with world-state-1 in the sense that an understander has the alternative of answering questions about world-state-1 by directly observing the world state or by questioning the corresponding knowledge state."

In the first quotation the thing which is represented is some abstract entity, "knowledge", and the way it is represented is by giving it a formal, theoretical structure which is sufficiently detailed and unambiguous to allow some inferences or computations whose results can be checked by observation. In the second quotation, the thing which is represented is a concrete situation, the state of the world at a particular instant, and the way it is represented is by associating it with another concrete phenomenon, called the knowledge state of a particular cognitive system.

This ambiguity of the word representation here is the same as the one we encountered in the context of physical science: there is one formal, or mathematical sense where representation means that you give some efficiently manipulable structure to an abstract concept (e.g. "knowledge" or a "group"), and there is a concrete sense where representation means that you construct some model or image of a concrete phenomenon (e.g. an external situation, or the movement of a falling body) (this ambiguity in meaning is similar to the one

encountered for the word "model" which has a quite different formal meaning in formal logic, compared to its ordinary use in science).

Apart from this difficulty, the representation concept in cognitive science is clearly related to its counterpart in physical science, though the technical meaning is not quite the same. This relation will be further elucidated by the analysis of a last specialized sense in which the concept of representation is used in cognitive science.

1.5 AI and problem representation

In the first computer systems there was no intention to provide the system with a perception or understanding of the outside situation. Rather the purpose of the first computer scientists was to construct systems which could efficiently *solve* complex *problems* which were introduced by the programmer, and had no direct relation to the concrete environment in which the system was working. E.g. if the programmer asked the system to compute 9231: 17, the system was not supposed to understand what those numbers meant, what they were referring to, but just to bring forth a new number which was the solution to the proposed problem.

At first these problem solutions were carried out by a programmed set of *functions* or *algorithms*, i.e. sequences of elementary operations on data or information units stored in memory which would lead unambiguously from the initial data (9231: 17) to the final result (543).

In a later stage one became aware that for more complex problems there was no determined sequence or path of operations which led directly from the given data to the result. Algorithmic methods had then to be replaced by methods of *heuristic search*: since for each step in the problem solving process different sequences of operations could be attempted, the system, to be efficient, should be provided with a set of rules (heuristics) which specified which path of operations should be tried first. There was however no guarantee that the application of the heuristic rules would lead to any solution at all. Therefore problem solving could be viewed as a search, guided by heuristic principles or rules-of-thumb, through a space of potential solutions or problem-states (Nilsson, 1971).

Here again we encounter the problem of complexity: if the space to be searched through is very large (possibly infinite) it may well be that the computer with its limited information-processing capacity, will never come up with the required solution within a practical time interval. This is where the peculiar limitations of computer science come in: the difference between mathematics and computer science is that a mathematician may well be satisfied when he has proven that a certain problem has a solution, whereas the computer scientist will not be satisfied until he knows some explicit procedure which will help him to effectively find this solution within a finite time. To a computer scientist, for a theoretical model to be meaningful it must specify how the problems it is supposed to describe can be practically solved through a finite sequence of

operations. (A somewhat similar philosophy lies at the base of "Constructive Mathematics", cf. Beeson, 1985).

This requirement leads directly to the problem of "representation" in a sense similar to the one mentioned earlier of "indexation". Indeed the main problem of a computer scientist who wishes to design a practical program for solving efficiently a certain class of problems, is to translate the abstract conceptualization of this problem class (e.g. multiplying integers, or proving theorems in the predicate calculus) into a detailed, indexed program or information structure, such that the problem solving process can be carried out in a minimum number of steps and with a minimum of search required at each decision point. This means in practice that one must introduce:

- first, a clearly indexed *space of states*, such that as well the given data as the looked for solution belong to that space;
- second, a set of *operators* which can be used to transform one state into another one; this set of operators must be sufficient to be able to reach the complete space without having to apply too many operators, but it must not be so detailed that it would take an excessively long time to decide which operator to use;
- third, a set of *heuristic rules* or *evaluation criteria* which give some indication about whether the chosen search path is likely to bring one closer to the solution; in particular the criteria must allow us to test whether the state reached is indeed a solution (or close to a solution).

A problem which can be represented in this way is called a "*well-structured*" or "*well-defined*" problem. In order to solve such a problem it suffices to carry out a search process through the state space, starting from the initial state corresponding to the given data, and then applying the different operators, guided by the heuristic rules, until a solution is found. The efficiency of the search will of course greatly depend on features such as the size and the structure (the way its points are connected through pointers, or geometrical relations) of the state space, the ease of application of the operators, the reliability, simplicity and exhaustiveness of the heuristic rules, etc. Here we see that the way a problem is represented is essentially important with respect to its practical solvability.

This becomes even clearer if we consider so-called "*ill-structured*" problems, where no set of states or operators is given as yet. The first step to solve such a problem is to design some sort of representation, which will allow you to start a search process. This step is crucial, and the way you carry it out will determine whether you are able to find a solution in any reasonable time at all.

A classic example is the "mutilated checkerboard problem". Suppose you have a checkerboard consisting of 10 × 10 black and white squares. Suppose the two white squares at the extremities of the diagonal line are deleted. The problem is now stated as follows: can you cover the remaining mutilated board completely

with domino blocks, knowing that one domino block covers just two adjacent squares? At first the problem seems simple and well-defined: the possible states of the space are the subsets of the board which are covered, the operators consist simply in putting a domino block so that two more squares are covered. It appears that you just have to try some combination of covering operators until you find one which covers the whole.

However, you will never succeed in finding this combination: whatever you try, there will always remain certain isolated squares that cannot be covered without uncovering other squares. The reason for this is seen very simply when the problem is represented in another way: instead of looking at an operator as just the covering of two squares, you should see it as the covering of two adjacent squares, which means of one black and one white square. Now since you cannot cover a black square without covering a white square, it becomes clear why a checkerboard where there are less white than black squares can never be covered in this way.

This is an example where a change of representation makes a simple-looking problem clearly insoluble. In other cases a change of representation can transform an at first sight very complicated problem into a simple one. Many more examples of this principle can be found in books and papers on problem-solving (Nilsson, 1971; Newell and Simon, 1975; Wickelgren, 1974; Amarel, 1968; Burghgraeve, 1976).

Once again we must ask how this concept of problem "representation" is related to the other representation concepts we already encountered. On the formal level there is clearly a similarity between these "problem representations" and the "dynamical representations" used for modelling systems. Both use the formal structure of a state space, and of a trajectory through this space determined by dynamical laws, or by heuristic rules.

But this similarity is more than purely formal. In systems theory one can use two equivalent descriptions of a system: *input-output systems* and *decision-making or goal-seeking systems*. (Mesarovic and Takahara, 1975). In decision-making or goal-seeking systems, the dynamical mechanism which allows us to infer output behavior from input data is supposed to be determined by a set of internal criteria that the system tries to satisfy or to optimize. The dynamical relation between input x and output y can then be defined as follows: a system S will react to the input x with the output y if y is the solution of a problem specified by the initial data x, and the internal goal or evaluation criteria of the system.

In practice both dynamical and goal-seeking descriptions are equivalent. This is easily seen in classical mechanics where the trajectory of a moving object can be described alternatively as the causal progression of a point in phase space, continuously determined by the dynamical law, or as the solution of a Hamilton optimization problem (Mesarovic and Takahara, 1975).

That in practice the concepts and structures introduced for constructing problem representations are very similar to the ones used in representations of physical systems is nicely illustrated by the formalism proposed by Amarel

(1968), for the representation of problems of reasoning about actions. We will show later, by carefully analyzing the categories and organizing principles used for constructing representations, that problem representations and dynamical system representations are fundamentally equivalent.

Now, what is the relation between problem and knowledge representations? In a certain sense a problem representation is just a special case of a knowledge representation: it represents the specific knowledge needed to solve efficiently a certain problem (class). In practice the difference between both representation concepts is often one of emphasis: in knowledge representation the emphasis is on the processing of incoming stimuli (understanding, perception) so that an eventual decision concerning this information can be made; in problem representations, the information is generally already pre-processed (e.g. by the programmer), the emphasis is on the solution of a particular problem determined by this information.

We now come to the point where we have found three more or less elaborated significations of the word "representation": *dynamical representations*, *knowledge representations*, and *problem representations*. Some other significations we have encountered (e.g. ideas as representations of outside phenomena, or matrices as representations of abstract operators), have no sufficiently broad domain of applicability to be useful on their own in tackling the problem we have formulated. Besides they can be subsumed under these more general headings.

We can now try to synthesize these three concepts into one, and to apply it to the problem of change. This synthetic concept will be called "adaptive representation" or "representation" for short, and will be discussed in the following chapter. The various applications of the representation concept we have discussed seem to indicate that this concept is very powerful. But in order to use it efficiently, we must first remove some of the remaining ambiguities and confusions.

2.1 Representations as an interface between Mind and Nature

Until now our discussion of the various uses of the representation concept has presupposed the basic Cartesian dualism between Mind, as studied by cognitive science, and Nature, as studied by physical science. However, this has led to a basic ambiguity as to what a representation is supposed to represent: the formal structure that was called a representation could be said either to refer to an abstract, mental entity (e.g. a problem, an algebraic concept, a piece of knowledge), or to a concrete physical phenomenon (e.g. an object, a physical process, a situation).

If we borrow some terminology from logic, we might call the first, abstract reference the "*intension*" of the representation, and the second, concrete reference its "*extension*". If we consider a representation as a special case of a "sign", we can also use the terminology of Peircean semiotics (see e.g. Hawkes, 1977), and call these two references respectively the "interpretant" and the "denotatum" of the representation. So, any general representation has both an extension, i.e. a set of outside, "objective", physical phenomena that it denotes, and an intension, i.e. an internal, "subjective", mental meaning or interpretation.

Most confusions with respect to the representation concept arise because only one of these two basic functions is acknowledged. In order to get a real insight in the power of this concept you must look at it from both viewpoints.

For example, if you have a mathematical description of a state space together with a set of dynamical operators and equations that allow you to compute output functions from input functions, then you can either look at it as a mathematical representation of a physical system (e.g. an electronic circuit), or as a representation of the knowledge that you have got about this system. (In fact it is also a representation of the knowledge that you have not got but might want to get by solving the problems inherent in the equations).

Therefore, a representation belongs neither to the realm of matter, of outside objects, of things-in-themselves, nor to the realm of pure mind or Platonic Ideas: it constitutes an interface, it stands in between Mind and Nature, in between *subject* and *object*, in between *Self* and *World*. You could go even further and say that the concept of representation transcends the classical dichotomy of Mind and Nature: the only things we have got to work with are representations; neither pure ideas nor things-in-themselves exist in any operational sense, they are by definition unreachable ideals.

In a certain sense the subject-object dichotomy is an artifact of representation: it is the representation itself which creates a distinction between "inside self" and "outside world" (De Mey, 1982; see further sect. 9.4). In general a representation can be seen as a mediating system which relates different domains. The distinctions made within these domains, the way their

parts are labeled as either subjects or objects, are determined by the structure of the representation. This means that, depending on the context, the same event can be represented either as belonging to the internal self or as belonging to the outside world. De Mey (1982) discusses two simple examples:

> When I reach out to touch a pot of coffee in order to find out if the drink is still warm, the same experience might be framed in either the expression *the coffee is hot*, or the expression *I burned my fingertips*, i.e. two different conceptual representations of the same bodily experience. One refers to an object in the outside world, the coffee, the other to an object which belongs to the self, my body. The common mistake is to think of the cognitive view as applying only to models of the external world. Representation, however, is prior to segmentation in self and world.

The other example is that of the Copernican revolution in the scientific representation of the movement of the planets: in the ancient Ptolemaic representation the apparent motion of the sun was considered as something belonging to the external world, to the object "sun" itself; in the Copernican view, however, the motion was ascribed to the Earth and to the subjects which stood upon it and looked at the sun.

So, at least in principle the representation concept appears to bridge the gap between the sciences of Mind and the sciences of Nature. We will now show that also in practice the need is felt in both physical and cognitive science to bridge this gap by using concepts related to "representation".

The great conceptual revolution that took place in physics in the first half of this century was centered around two theories: *relativity theory* and *quantum mechanics*. In relativity theory the main new insight was that basic physical properties such as space, time, mass and energy, are dependent upon the reference frame of the observer. This reference frame can be seen as a special case of a representation (see sect. 8.1).

In quantum mechanics too the observer using a representation plays a predominant role. From the indeterminacy principle it follows that there is always a finite interaction between the subject and the object he observes; this induces a perturbation of the "object-in-itself": there is always a remaining indeterminacy, due to the non-separability of subject and object. Since the observer cannot have a complete knowledge of the object, he must choose that representation of the state of the object which is most useful for computing the variables he is interested in (see sect. 7.1).

For example, if he is interested in the probability of finding a particle in a certain spatial domain, he will use the Schrödinger or wave function representation of the particle state. If he is interested in the possible values of the particle's spin, he will represent the state as a linear combination of spin eigenvectors. The peculiarity of the observation process is that an object which is in an eigenstate of a particular representation (i.e. the observable properties—e.g. spin—corresponding to this particular representation have determined values) will in general no longer be in an eigenstate of this representation after

an observation is made, but in an eigenstate of a new representation corresponding to the property which was observed (e.g. position).

This strange effect suggests that in quantum mechanics representations are more than essentially equivalent and purely conventional descriptions of an invariant physical phenomenon. It appears that subject, object, representation and observation process can no longer be treated separately: they form one, indivisible whole (cf. Finkelstein, 1979).

Yet, the traditional physical world view lacks the conceptual tools required for modelling this basic interactivity. This leads to the appearance of a host of paradoxes and conceptual problems associated with quantum theory (see e.g. Jammer, 1974). During the last few decades the awareness has been growing that the only way to solve these problems is to concentrate not so much upon the physical mechanisms underlying quantum phenomena, but rather upon the epistemological (i.e. cognitive) aspects of physical observation and representation. One important approach which grew out of this trend is the analysis of the logical structure of theories or representations based on different types of observations (see e.g. Piron, 1976; Aerts, 1983).

A complementary evolution has taken place in the cognitive sciences. At first it seemed fruitful to study mental or cognitive phenomena in isolation, without reference to any outside physical world. For example one would analyze the abstract structure and properties of logical formalisms, of generative grammars, or of general heuristics for state-space search.

Later on, it was acknowledged that in order to model practical intelligence, you should study systems capable of facing real-world problems, of interacting with concrete situations. One of the domains where this need is felt most clearly is robotics: an intelligent robot should be able to experience the material world, to act upon it according to its proper goals, and to correct its actions by adequately interpreting the feedback it gets.

On a more theoretical level, it was proposed to study the representation of so-called naive or qualitative physical knowledge, that is to say the knowledge people use for their everyday interaction with the physical world (Hayes, 1978; De Kleer and Brown, 1983; Forbus, 1981). This would provide a basis for constructing AI-programs, which would extend beyond the traditional "toy problems".

These examples illustrate that in cognitive science too there is a trend towards a more holistic or interactive approach, in which both the knowing subject and the known object are encompassed. The purpose of the present work is to carry this trend to its logical endpoint, and to study the domain where the physical and cognitive sciences meet. Therefore we must define an integrative concept of representation, and establish its relation with the fundamental problem of change.

2.2 Adaptation as vicarious selection

The representation concept we wish to define should be a characterization of the dynamical relationship between subject and object, between mind and nature, between the inner and the outer environments of an intelligent, adaptive system. We shall suppose that a "subject" or "self" is characterized by a *stable identity*, and that it continuously tries to maintain this identity by reacting in an adequate way to all (external) perturbations that might destroy it. Hence, the fundamental purpose of the "self" system is survival: maintenance of its identity.

From this viewpoint all interactions between such a system and its environment can be reduced to processes of *selection*: either the system survives the interaction, i.e. it is positively selected, or it is destroyed, i.e. it is negatively selected. So, the role of the environment can be conceived as that of a dynamic selector, which continuously eliminates certain systems while retaining other systems.

This description is so general as to become tautological: in any complex process of change it is possible to distinguish phenomena that are stable, i.e. which maintain some kind of identity during the interval of observation, from phenomena that are unstable, i.e. which are destroyed during this interval. The unstable phenomena will be considered as a background (or environment) of change against which the stable phenomena stand out. As will be discussed later, the stable phenomena can be categorized in a hierarchical fashion, according to the way they maintain their stability.

We shall call a high-level stable phenomenon an "*adaptive system*" if it can *anticipate* in some way the changes to which it will be subjected, so that it is able to prepare its defense against potentially destructive processes.

A basic example of such a system is a human being. If a person is confronted with a potentially dangerous phenomenon (e.g. a fire), he does not have to experience the destructive influence of the phenomenon (e.g. being burned) in order to take counteraction (e.g. running away); it suffices that he perceive certain warning signals (e.g. smoke) in order to infer the danger to be expected, and to conceive of an appropriate plan of action. In order to be able to make such an anticipation, the system should have some form of knowledge of the world (e.g. that smoke is an indication of fire) and of its self (e.g. that it can be killed by fire).

The simplest way to conceptualize such knowledge is by considering it as a "*vicarious selector*" (Campbell, 1974): it selects those possible actions of the system, which lead to its survival in a given external environment. This selection is vicarious because it "stands for", i.e. represents, the selection effectuated by the environment itself through the destruction of those systems that behave in an inappropriate way.

Let us illustrate this principle by working out the fire example. Suppose that different organisms (e.g. people and animals) are confronted with the same situation of an approaching fire. Suppose that they react in different ways: some of them run away, some hide, some merely continue the things they are doing. After the fire has passed, a selection has taken place: only those organisms have

survived who reacted in an appropriate way, e.g. by hiding in a fire-free place, or running away far enough. So, the fire can be considered as a selector of adequately behaving systems.

Consider now an individual. Let us assume for simplicity that just two cases can be distinguished: either he knows how to react adequately when perceiving the signals of fire, or he does not know. In the first case he will select the appropriate behavior from his repertoire of possible actions, and survive. In the second case he will make a "blind guess", and adopt a certain behavior without knowing whether it is appropriate or not. In that case he will have to wait for the continuation of the events to see whether the choice was adequate or not.

So, we may conceive of knowledge as something that allows us to select appropriate behavior vicariously, before the environment does so by destroying (or damaging) the inadequately behaving systems. This concept of a vicarious selector can be seen as a first step towards the integrative concept of representation we are looking for. The fundamental idea is that a system can succeed in maintaining its stable identity in spite of the continuously changing interactions with its environment by internalizing the "selective action" of these interactions upon its behavior, so that it can anticipate which type of behavior will be adapted to a given situation. In order to do that it should at least be able to *distinguish* between interactions that are dangerous (i.e. which select negatively) and interactions that are not.

The question we must answer now is: how does such a vicarious selector "internalize" or "represent" the selectivity of interaction with the environment?

2.3 The thermostat as an example of an adaptive representation

We may clarify this problem by considering an example of what De Mey (1982) calls "the simplest model of a world view (or representation)": the thermostat. Suppose that you have got a tank with tropical fish, and that in order for the fish to survive, the temperature of the water must be kept above 24° C. The apparatus that is normally used to maintain such a system in equilibrium consists of a heating element connected to a thermostat, both installed inside the tank.

The thermostat consists of two parallel metallic plates: one is fixed, the other one can move, and may or may not come in contact with the fixed one. The movable plate is bi-metallic, which means that it will react to a temperature decrease by bending towards the fixed plate. If the temperature falls below some threshold (e.g. 24° C), the two plates make contact. This closes an electric circuit which activates the heating element, so that the temperature is raised again. If the temperature becomes high enough, the bi-metallic plate bends back again and the contact is broken. This stops the heating activity so that the temperature can decline again.

This is a classic example of a *feedback* system. But it is also an example of an elementary adaptive representation. We have indeed a system with two states: plates in contact, and plates not in contact. The first state of the thermostat corresponds to the state of the world:

"The temperature in the tank is too low (below 24° C)",

the second one to the world-state:

"The temperature in the tank is high enough (above 24° C)".

In that sense the thermostat can be said to have an elementary knowledge of the world around him: it knows whether the temperature is below or above some threshold. But it also knows about the adaptive system to which it belongs, namely that the temperature should ideally be 24° C for the system to maintain its equilibrium.

This "*self-knowledge*" is embodied by the fixed plate of the thermostat, which forms the standard against which the movements of the bi-metallic plate caused by temperature changes are evaluated. The position of the bi-metallic plate then is an embodiment of the knowledge about the present interaction between system and environment, which is arrived at through an elementary form of perception. So, the subsequent states of the thermostat represent as well the changes in the external environment (temperature fluctuations), as the stable identity of the system (its equilibrium temperature).

The thermostat is also a vicarious selector: it selects the right temperature for the fish tank. Suppose you would not have a thermostat determining the right temperature, and you would keep different, but equivalent fish tanks at different temperatures (e.g. 16° C, 19° C, 24° C, 34° C, …). In that case the selection would not be vicarious but direct (i.e. natural): most populations of fishes would die, except those kept at a temperature close to 24° C.

Let us analyze which basic processes and structures (i.e. aspects of change, respectively of stability) are governing this system. On the stable side we have the structure of the thermostat: the fixed plate, the moving plate which for a constant temperature is at a determined distance of the fixed plate, the bending constant of the moving plate which is such that the distance will go to zero when the temperature goes to 24° C. This structure is a material realization or representation of the abstract, "mental" organization of the system: its goal of remaining at a temperature near 24° C, its ability to perceive or distinguish whether the temperature is above or below this threshold, its knowledge that to reach the goal the heating should be activated whenever the temperature falls below 24° C, and should be deactivated whenever the temperature is above 24° C, its ability to act upon the situation by opening or closing the heating circuit.

In fact we find here, in its most primitive form, the fundamental qualities characterizing "mental" activity: purpose or *motivation*, *perception*, *knowledge*, *problem-solving*, and *action*. Yet we know that to construct such a system it is sufficient to have an elementary knowledge of physics: the relationship between temperature and the differential expansion of metals, the warmth dissipation through a closed electric circuit. In this sense this extremely simple system is on

the boundary of physical and cognitive science: it is basically governed by physical principles, but it adds to this something that is outside the realm of pure physical theory: a very primitive "mental" or cybernetic organization.

Let us now look at the dynamic aspect of the phenomenon: what processes is it going through? Let us begin with perception: a change in the outside temperature causes a change in the bending of the bi-metallic plate. We may say that the system has become "aware" of the temperature change, it has perceived it. Now this perception is evaluated: is the temperature below (contact between the plates) or above (no contact) 24° C? After evaluation comes a phase of inference (cognitive processes or thought): if the temperature is below, then the heating should be activated, but if the temperature is above, then it should be turned off. Finally there is the phase of action: the "message" is sent to the heating element that it should or should not heat.

Suppose the element begins to heat. This adds warmth to the water of the tank and changes its temperature. The lower the outside temperature, the slower the inside temperature increases. After a certain time interval, which depends on these outside contingencies, the inside temperature increases above the threshold, and a new perception-evaluation-inference-action sequence is triggered. So, the process proceeds in a circular fashion. This is called a *feedback loop*: the external consequences (temperature change) of the systems action (output) is fed back into the system (input) and re-evaluated so that a new, corrective action can be initiated.

In fact there is also another aspect in the process which can be called prediction or "*feedforward*". It means that the action is initiated not only by outside reactions which determine the direction in which the previous action should be corrected, but also by the anticipation that a particular action will bring the system closer to its goal independently of any outside contingencies. In general an adaptive system will as well use feedback as feedforward mechanisms in order to deal with changes in the outside environment (cf. Simon, 1979, p. 172).

In the thermostat example the feedforward mechanism is very primitive: the only anticipation made is that if the temperature is below the threshold, then the activation of the heating element will bring the system closer to its goal. This is because the representation used by the system has only two states: its only internal knowledge is that a certain action will transform one state into the other one.

2.4 The interdependence of feedback and feedforward

If the representation becomes more complex (i.e. has more states) then the possibilities for anticipative knowledge, and hence for feedforward, become much larger. Between a given state and the goal state there will now be many paths of action, and the system will have to decide which sequence of actions is most probable to lead to the desired goal state. In this case the set of states-of-awareness will get a much more complex structure, which will allow the system

to make more detailed and long-term anticipations, so that it can adapt to more complex and changeful environments.

Yet it will always need a feedback mechanism to correct, or to complete inadequate anticipations, because a slight error or incompleteness in the representation of a situation can be greatly enlarged by the active application of feedforward, so that the system loses its stability (cf. Simon, 1979, p. 173). This interdependence of feedback and feedforward is crucial to the working of adaptive representations.

Let us illustrate this principle with an example. Suppose that you are exploring a region, with only a map and a magnetic compass to guide you. By comparing the features of the landscape with the contours on the map you have succeeded in locating your position with respect to the map. This provides you with a representation of your initial state. Suppose you now wish to reach a mountain hut which would provide a shelter for the night, and that you look it up in the map. The symbol on the map that indicates the location of the hut is a representation of your goal state. You can now draw a line on the map that connects your actual and your desired position. This line represents your plan of action, it anticipates the path you are going to follow in order to reach the mountain hut. In principle, the problem now appears to be solved: it suffices that you consequently follow the direction indicated by the line on the map to be sure to reach the hut. Until now, the only adaptive mechanism you have used is feedforward: with the help of the representation provided by the map you have simply anticipated the different actions that you have to perform in order to reach your goal.

However, anyone who has done this type of exploration in reality knows that this is not sufficient. Even though you may use a compass to guide your steps, so as to be sure that you continuously follow the direction indicated by the line on the map, you will always deviate somewhat from the ideal path. This may be due to different factors: your march is always somewhat irregular so that it is impossible to walk in a perfectly straight line; you may encounter obstacles on the road (e.g. rocks, bushes, pools, …) which were not indicated on the map, either because they were too unimportant to be included in the detail of the map, or because the terrain has changed since the map was drawn (e.g. because of an avalanche, or the movement of a river bed).

Unless this deviation is corrected, it will become larger and larger, until you have lost any idea about your position with respect to the anticipated path. The only way to tackle this problem is by using feedback, that is to say by regularly comparing your actual position with the anticipated position as indicated on the map, and by redesigning a corrected path which takes into account the eventual difference between both positions.

This method of correction is very general, and the more frequently you apply it, the smaller the risk of failing to reach your goal. The only presupposition you must make is that of continuity, i.e. that a "small" action will have only a "small" effect on your state, so that the deviation between actual

and ideal state will not grow in a discontinuous fashion. However, this is not necessarily so. For example, taking one step too much might result in falling into a precipice, and this would probably eliminate all chances of ever attaining your goal. The only way to handle this kind of discontinuous changes in the environment is by using feedforward or anticipation, that is to say by using a good map on which all precipices and other potentially dangerous phenomena are indicated.

Let us conclude this discussion of examples by remarking that both feedforward and feedback as adaptive mechanisms for coping with change are fundamentally dependent on the representation used by the adaptive system. This fact is ignored in most presentations of the feedback-feedforward concepts, but it is essential for the understanding of change and adaptation. It was made clear by the exploration example that anticipation or planning is heavily dependent on the map (i.e. the representation of the environment) which is used.

But also the corrective action through feedback requires the use of the map, in order to recognize how far, and in what direction the actual path has deviated from the planned path. You cannot take corrective action if you do not know how to interpret the reactions or signals you get from the environment. The discussion of knowledge representations has made clear that this interpretation process (which encompasses perception, inference and evaluation) cannot be carried out effectively without a representation. This explains why even the most primitive cybernetic systems, such as thermostats, should incorporate some representation of their goal state and of the possible states of the environment in order to be efficient.

2.5 Structures and states of an adaptive representation

Now that we have discussed different examples of representations in order to get a feeling of what this concept means in practice, it is time to integrate the different features we have found into a general, abstract model. First we may characterize an adaptive representation as an abstract organization, which can be realized or embodied in different concrete media (e.g. electronic circuits or nervous tissues). These material realizations are not fundamental to the representation, and will therefore not be discussed.

The function of the representation is to *steer* or to *guide* the interactions between a system and its environment in such a way that the "identity" of the system is maintained throughout the changes occurring within the environment. This allows the system to adapt, that is to say to change internally in such a way that the external changes are compensated before they can destroy the identity.

In order to do this, the representation should provide a model of the possible changes in the environment, of the possible actions of the system, and of the identity or "self" to be maintained (this differentiation between features of the world (external) and of the self (internal) is determined by the structure of the representation, it is not determined a priori). This allows the system to explore the environment (or rather the possible relations between system and

environment) vicariously, i.e. to test possible strategies for behavior with respect to the external situations without actually carrying them out, and hence without running the risk of being harmed because they were inadequate.

The relation of "vicariousness" between environment and representation is not a static mapping or *isomorphism*, in the sense that a photographic picture is the (static) result of a mapping of certain (visual) aspects of a particular environment. Most people who think about a "representation" tend to conceive of it as something that has a formal resemblance to the thing it is supposed to represent. The correspondence between environment and adaptive representation, however, is not one of structure or form (static) but one of process or evolution (dynamic). This means that in general there will be no correspondence between parts (substructures) of the environment, and parts of the representation.

The idea that a representation should result from the mapping of parts or elements of an environment onto the elements of an abstract structure, will be called the "*denotative*" or "*isomorphic*" view of representations. It corresponds to the view that the meaningfulness of a linguistic description results from the fact that the different words of the description denote different parts of the phenomenon that is described, or that the meaning of a picture resides in the isomorphism between its parts and features, and those of the thing it is supposed to represent.

This mechanism of representation corresponds to what De Mey (1982) calls the "monadic" and "structural" stages of information processing (see sect. 1.4), where a stimulus to be processed (or a situation to be represented) is decomposed into separate units which are related through invariant structural features. In the later stages (which De Mey calls "contextual", respectively "cognitive"), however, it becomes clear that the interpretation of a situation generally requires more than an analysis of its (static) structure, it requires knowledge about how this situation has evolved, and how it may evolve further on.

It is this emphasis on the dynamics of environments or situations that makes a representation "adaptive". In this sense a painting representing a still life, or a poem describing a love affair, are no *adaptive* representations, they are static representations.

Now, what makes a representation *adaptive*, what mechanism allows it to model actual and potential *changes*? This mechanism is based on the duality between *state* and *structure*. The structure is the stable part of an adaptive representation. It is an organization or interconnection of elements that does not change (or changes only slowly with respect to the information processes that it directs). The state is the aspect of a representation which changes continuously, in correspondence with the changes in the environment it is supposed to represent.

The transition from one state to another one is determined partly by the representation structure (which specifies the possible states and the possible transitions from one state to another one) and by the actual outside situation as

perceived by the receptors of the system (which singles out one of the potential states or transitions). For simplicity, we shall suppose for the moment that the distinction between state and structure is sharp, although we shall see later that for certain representations (e.g. neurological networks) the distinction is only gradual, in a way similar to the distinction between short-term and long-term memory.

As an example, consider a language, characterized by a vocabulary or lexicon, a grammar or syntax, and an internal logic. Lexicon, syntax and logic correspond to the structure of the representation, which must be known in order to use it effectively. A language user who knows this structure possesses what Chomsky calls "*competence*": he is able to produce well-formed sentences.

The sentences he will produce, however, are not completely determined by this structure, but will depend upon his intentions and the situation he is in. These actually produced sentences are what Chomsky calls the "*performance*" of the speaker; they correspond to the state of the language system viewed as a representation. (A similar correspondence exists between de Saussurean concepts of "langue" respectively "parole", and the concepts of "structure", respectively "state" of a representation).

The flexibility of a language system results from the fact that the number of possible sentences (i.e. states) is unlimited, whereas the number of words and logical-syntactical rules (i.e. structures) is finite. This allows the language user to adapt his verbal utterances to describe the most diverse situations. The only thing which is missing in this characterization of a language to make it a full-fledged adaptive representation, is the aspect of dynamics: the transition from one linguistic utterance to another one, is not specified within the language, but is dependent upon the user.

This example gives us some clues about what is needed to build a representation structure. First, a representation structure must provide some *generative mechanism* for producing different states. As in a language, a state can be conceived as a combination (e.g. a sentence), determined by specific constraints (e.g. syntactical rules) of units (e.g. words), chosen from a predetermined set (e.g. a lexicon). The set of allowed states, together with their interrelations, forms a *state space.* The set of units will be called the unit space. In the trivial case where the maximum number of units to be combined is one, state space and unit space coincide. In the case where the number of units that may be combined in a single state is unlimited, the state space will be infinite, even though the unit space may be finite.

The selection of one combination of units from the set of all possible combinations, can be conceived as an "*actualization*" or "*activation*" of the chosen units. Hence, the situation where a representation is in a particular state can be viewed as a *selective activation* of units from the representation structure.

Beside the constraints which determine the allowed or potential states (generative constraints), there must be constraints which determine the allowed

transitions between states. This may be called the *dynamical* part of the structure.

The third type of constraints determines which transitions are in accordance with the goal (self-maintenance) of the system. These may be called *purposive* constraints.

It should be noted that this subdivision of representation structures into unit space, generative constraints, dynamical constraints and purposive constraints is purely conceptual: there is no reason why these different aspects of the functioning of an adaptive representation structure should also correspond to actually distinct substructures. As we shall see (chapter 4), the explicit representation structures exemplified by scientific models are indeed constructed according to this pattern, but this is no longer obvious for the implicit representations exemplified by neural systems (e.g. human brain systems).

We should also remark that the things which were called static representations, such as a spoken utterance, a painting or a poem, can be viewed as (externalized) states of an adaptive representation, whose structure is determined by the skill or competence of the one who created them. This skill needed for constructing adapted static representations is sometimes called "mode of representation" (Blatt, 1984) or "representation technology" (Cohen, 1979; 1982). In the present terminology, these terms are equivalent to "representation structure"

2.6 Information-processing in an adaptive representation

The interactions between the stable structure and the changing environment can be subdivided into *input*, i.e. changes of the representation state resulting from the external situation, and *output*, i.e. changes of the external situation resulting from the representation state. In a higher-order cognitive system the input process corresponds to *perception*, the output process to *action*. The process which leads from input to output is determined by the representation, and can be called "feedforward". The process which leads back from the output through a change in the environment to a new input, is called "feedback".

The interaction process as a whole can be conceived as a loop, leading from the representation, through the environment, back to the representation. The process, however, is not periodic, since each time the cycle has been completed, the representation is in a different state.

This process can be further conceptualized as a transmission and processing of some abstract substance, which may be called "*information*". The information concept is confuse, and has many different significations (Nauta, 1972). The signification we will use here is very simple, but general. The basic idea is that the acquirement of information implies the selection or distinction of a subset of actual or probable cases from a larger set of potential cases.

For example, if I throw a coin, there are two possible outcomes to be expected, but I do not know yet which one will show up: I lack information. Now, when I look at the result, and see that the coin has fallen with the face up, I have received one bit of information: the actual event (face up) has been selected from a set of potential events (face up, or face down).

The quantitative measurement of information (bit measure) is studied by information theory, and is based on the probability measures of the different potential events from which a subset is selected (Nauta, 1972; see further sect. 9.2). In practice, it is very difficult to introduce such quantitative measures, and we will use the information concept primarily in a qualitative fashion.

In the sense we will apply it, any process of selection whereby a distinction is made between phenomena which are retained, and phenomena which are eliminated, is basically a transfer of information from the selecting agent to the set of phenomena to be selected. The set of retained phenomena (e.g. events, systems, or states) can be said to incorporate or to store the transferred information.

Let us apply this scheme to the interaction between environment and representation. The input stage can be viewed as a selection of potential states of the representation by the environment: only a part of the potentially activated elements of the representation are actually activated by the signals passing through the perception apparatus. Hence, the representation has received information about the environment: it now knows which of its potential states is an acceptable representation of the actual outside situation.

However, the adaptive representation is not merely undergoing the selective influences of the environment in a passive way, it is itself actively selecting the information-carrying signals it receives. This internal selection is carried out by the knowledge embodied by the representation (see sect. 1.4).

This "knowledge" can be seen as "stored information", whereas "information" can be seen as "transmitted knowledge". So, the knowledge inherent in the representation structures selects, and hence, in a certain sense, "adds" information to the passing signals from the environment. This phenomenon is called "*information processing*": the information is transformed, partly through the loss of information which was filtered out, or has dissipated spontaneously, partly through the addition of new information coming from the processing structure. The "end product" of this representational information processing may be called "meaning".

This can be conceived as follows: the (selective) influence, or signal, coming from the environment activates a certain state of the representation; this state embodies the received information; it will undergo an internal evolution (information processing), determined by the constraints of the representation structure until a (provisional) "equilibrium" state is reached (i.e. a state which does not have to be processed further); this state represents what the signal "means" with respect to the representation and its inherent goal; from this point on, a new process can be set in, which uses this interpretation or meaning of the

external signal in order to construct a plan designed to solve the problems defined by the discrepancy between the goal state of the system and the actual state resulting from the interpretation of the perceived signal; this plan can then be realized as a series of actions performed by the system through the use of effectors (e.g. muscles, or tools).

These actions, which constitute the output of the system, can still be viewed as information transfers from system to environment, because they select one out of several possible situations in the environment, namely this situation which is most likely to lead to the fulfillment of the systems goal. The information associated with this situation, which is partly determined by the system's action plan, partly by the reaction of the environment, can then be fed back into the system through the mediation of the receptors (i.e. sense organs), and so the cycle continues.

This sequential model of information processes is simplified in the sense that it neglects various intermediate feedback loops, which lead back from one stage of the information processing to a previous stage without passing through the complete input-system-output-environment-input cycle.

For example, there may be a feedback from the "equilibrium state" to the phase of immediate perception, because the "equilibrium state" is inconsistent with other informations, so that the perception stage must be checked and possibly corrected (e.g. you may hear someone speaking and reconstruct the sentence he spoke from the sounds you thought to hear, and then find out that this sentence is nonsense to you. In that case you normally direct your attention back to the sounds he uttered, and try to understand them in a different way).

These intermediate feedback loops contribute to the efficiency and stability of the overall information processing, but since they make the modelling often unnecessarily complex, we will neglect them most of the time.

Summarizing sections 2.5 and 2.6, we may say that a representation can simultaneously incorporate two types of knowledge (or information):

1) the stable knowledge inherent in the representation structure, which directs the processing of incoming signals and the planning of actions (in psychology this would correspond to the "*long-term memory*"),
2) the changing knowledge or awareness of the present situation, which is carried by the representation state, and undergoes continuous transformations determined by new perceptions and by internal processing (this corresponds to the "*short-term memory*").

We know from psychology that short-term memory or knowledge can be transformed to long-term knowledge. This process corresponds to a change of representation structure (cf. Korf, 1980), and is very difficult to describe since the model is based on the stability of this very structure. The only way to model this type of change is by transcending the state ↔ structure or short-term ↔ long-term dichotomies, and by seeing representation structures as the states of a

higher-order "metarepresentation". How this could be done will be discussed in some of the following chapters. But first we shall have to analyze some existing representation structures more explicitly, in order to uncover their implicit organizing principles, and to construct a detailed, formal model of a representation.

3.1 Scientific theories as explicit representations

We have described representations as the dynamic structures which form the backbone of knowledge. But how does *science* as an enterprise of knowledge fit into this framework? Clearly scientific theories and models are special cases of adaptive representations. But what distinguishes a "scientific" representation from a general knowledge representation, such as the one you or I use when ordering a dish in a restaurant?
For over a century, this problem has been the focus of philosophy of science. Unfortunately, there seems to be little agreement among the different philosophers of science about the right solution. However, the concept of representation provides a useful framework for the integration of the different approaches (cf. De Mey, 1982).

A simple classification of the activities of science might look as follows: observation, explanation, problem solving, prediction and testing. These functions correspond to the basic information processes occurring in an adaptive representation: a phenomenon is perceived or observed, an interpretation or explanation of this perception is generated; the problems arising from the discrepancy between the perceived situation and the desired situation are solved and a plan of action is conceived; this problem solution or action plan incorporates an anticipation or prediction of the future changes of the perceived situation; the actual changes are fed back through the perceptual system and compared with the predictions; if the prediction is not confirmed, a corrected interpretation and action scheme is generated.

The difference between these subjective, individual processes, and the social, *intersubjective* processes of scientific research, is that the latter processes must be made *explicit* in order to communicate them efficiently among the members of a scientific community. One way to exteriorize the structures and processes of internal knowledge representations is through the use of language as an external symbol system.

As was remarked earlier, however, language is not a complete adaptive representation, since it misses the dynamic, interactive features of such a representation: feedback and feedforward processes are not incorporated in the system of verbal language, but must be provided by the language user. Although language can be viewed as a social representation, i.e. a representation of the possible interactions or communications between two subjects (instead of subject and object), it is in itself not adaptive, it provides no goal or direction for the communicative processes which use it.

Therefore to have a social or intersubjective representation which is also adaptive, you must add to the language system, which determines the state space of the representation, a set of methods, procedures and evaluation criteria, which

can determine the direction or the trajectory of the processes occurring in the state space.

Such a system of values and procedures is normally provided by the culture in which the social interaction takes place. It is however seldom made explicit in the form of laws, rules or regulations. Moreover there is generally no explicit feedback mechanism: there are no intersubjective rules for checking whether the problem solution or action plan derived by feedforward from the social representation is indeed adequate with respect to the actual situation. As such most intersubjective representations proposed by social entities such as religion, culture, tradition, mythology, moral, etc., are incomplete and difficult to test or to correct.

Science can then be defined as the attempt to construct intersubjective, adaptive representations in which both feedforward and feedback mechanisms are made as explicit as possible. (It should be remarked here that the explicitation of problems and goals in science is only local or short-term; the global, long-term aim of science, the growth of knowledge, remains rather vague. In this respect science is less explicit than many other cultural systems, such as ethics, art, religion, and politics, which propose more general values and goals).

3.2 Formalization and paradigmatic structures

The methods used by science for making implicit representations more explicit, are *formalization* and *operationalization*. Formalization means that the representation structure is brought to the surface, so that "content" is changed into "form". In a perfectly formalized theory or representation there is no more hidden meaning or content to be uncovered, all relations between the elements of the representation are explicit and unambiguous. The classic example of a formalized representation (or formalism) is an *axiom system*. This consists of a set of semantic elements which can be combined into propositions through the application of syntactical rules. The semantic relations between the elements are specified by the axioms, i.e. a set of basic propositions which are supposed to be always true. The dynamics of the representation is specified by a set of deduction rules, which allow us to derive true propositions (theorems) from true propositions (theorems or axioms).

The meaning or content of a proposition is supposed to be completely determined by the set of axioms and deduction rules. Therefore, there should be no ambiguity about whether a proposition is true or not, if the axioms and rules are explicitly known (as shown by the Gödel theorem, and the lack of decision procedures for more general formalisms, however, this is not true in practice).

Outside of mathematics, this type of rigorous formalization of scientific theories is almost never achieved: there always remains some ambiguity about the meaning of the terms used to describe the object of study, and their interpretation requires a certain intuition. In general, the more complex and changing the domain of study, the more difficult it is to achieve a certain

formalization. Yet the scientist will always strive to define his terms rigorously and to state their interrelationships explicitly.

In so far as this attempt at formalization is successful, the intersubjective scientific representation becomes more or less well-defined or well-structured (see section 1.5). This means that there are definite procedures for solving the problems formulated within that representation. In general only a fraction of the problems considered relevant by the researcher are formulated and solved within the formalized part of the representation structure.

These explicit parts of the representation (well-defined concepts, rules, evaluation criteria, …) together with the examples of solved problems, form what Kuhn (1962) calls a "*paradigm*" (see also De Mey, 1982). The problems which are not yet solved but can be formulated within the representation, correspond to what Kuhn calls "*puzzles*". They will be more or less easy to solve, depending upon the degree of well-structuredness or formalization of the representation.

The already solved problems may function as prototypes or exemplars, from which heuristic rules can be derived by generalization or analogy. For example, in quantum theory the calculation of the energy levels of the hydrogen atom functions as an exemplar of a solved problem. More complex problems, such as the computation of the energy levels of the helium atom can then be solved through analogy with the hydrogen case. It is interesting to note that although quantum theory is considered a completely formalized and hence well-structured representation, it still needs exemplars such as the hydrogen atom or the one-dimensional harmonic oscillator in order to guide the search for problem solutions.

This should remind us that only for the most simple problem domains (e.g. addition or multiplication of integers) the sequence of operations needed to solve a problem can be completely determined beforehand, as a function or as an algorithm. In all other cases a certain amount of trial-and-error is necessary, and the less well-structured the representation, the larger this amount becomes.

We may conclude that the normal scientific models have the form of a paradigm, i.e. a partially explicit, intersubjective representation, in which the lack of general and definite rules for problem solving is compensated by the reliance upon exemplars which provide heuristic information.

Besides "*normal science*" Kuhn (1962) distinguishes two non-paradigmatic stages of scientific development: pre-paradigmatic science and revolutionary science. The former is said to occur before an intersubjective representation structure is established, the latter after this representation has become obsolete. This obsolescence is normally due to the fact that problems which belong to the domain of the representation (i.e. which are well-defined within this representation) cannot be solved, or that their proposed solutions are inconsistent with empirical facts or logical principles (this is called anomaly). In that case the need is felt for the replacement of the obsolete representation by a new, revolutionary one, but which is still to be constructed.

Both non-paradigmatic stages are characterized by a great confusion and lack of directionality or guiding principles. This is typical for the absence of a stable representation structure. In these stages scientific research will no longer consist of a search for the solution of well-structured problems (puzzles), within the representation, but of a search for a new, explicit representation of the problem domain.

The creative process of formulating a new representation is traditionally called "*discovery*" (see e.g. Simon, 1977). Since there is no overall set of explicit rules to guide this process, its only guides will be the implicit (i.e. intuitive) representations of the individual scientists, combined with some partial representations (i.e. exemplars, methodologies, concepts, …) taken over from other disciplines or problem domains. This explains why the discovery process is so difficult to understand, since no coherent representation seems to guide it. The modelling of discovery requires a more profound understanding of what a general representation is, and how it is constructed, i.e. how state and structure interact in its evolution.

3.3 Operationalization and empirical tests

The second characteristic which distinguishes scientific, explicit representations from ordinary, implicit representations, is operationalization. Whereas formalization aims to make the representation structure, and hence the feedforward mechanism, more explicit, operationalization aims to make the feedback mechanism more explicit. This means that the relation is specified between the states of the representation and the outside situations which they are supposed to represent, in such a way that the state which has evolved through feedforward can always be compared with the actual phenomenon to which it corresponds. For example, suppose that we have a representation of a physical system, which consists of a state space and a dynamical equation determining the possible trajectories of the state as a function of time. In principle, given a state s_1 at time t_1 we can then compute the state s_2 at a latter time t_2. State space and equation determine the representation structure; the computation of s_2 (t_2) from s_1 (t_1) is an application of the associated feedforward mechanism.

In order to apply this result to the concrete situation, and hence to check its correctness, however, we must specify how the symbolic entities s_1 (t_1) and s_2 (t_2) are related to the actual process occurring in the system. This means that we must establish a correspondence between s_1 (or s_2) and the results of a set of observations of the system at time t_1 (or t_2). If the same set of operations which yields the empirical values corresponding to s_1 at t_1, also yields the results corresponding to s_2 at t_2, we may conclude that the feedforward process is an adequate representation of the systems evolution between t_1 and t_2, and that no correction is needed.

The representation can be said to have been *confirmed* or (partially) verified. However, if the results of the empirical operations at t_2 do not correspond to the state s_2 as derived from the state s_1 observed at t_1, the representation has been *not confirmed* or falsified.

Therefore, operationalization of a representation means that the community of scientists working in a specific problem domain or representation agree upon a set of operations which allow us to establish the correspondence between the states (or propositions) of the representation and the phenomena to be represented. Following the input-output scheme we may classify these operations as either *detections*, or *preparations*.

Detection means that a phenomenon is observed or measured, and that the results of these observations (e.g. coordinates) are coded and mapped upon a state of the representation (e.g. a vector in a state space). This is an explicit or intersubjective counterpart of the process of perception.

Preparation means that a system is acted upon in such a way that it evolves to something corresponding to a given symbolic state. For example, a particle in an accelerator may be subjected to an electromagnetic field until it reaches a certain level of energy. It is then "prepared" in the state corresponding to this energy level. The operation of preparation is an explicit version of the action process.

The explicit version of the feedback process is called an "*experimental test*" of a representation, and is usually carried out according to the "preparation-detection scheme".

This means that a system is prepared so that it corresponds to a certain state s_1 (or the system is observed to be corresponding to a state s_1). Through feedforward the state s_2 is derived from s_1. (The relation between s_1 and s_2 may be static-logical or dynamic-causal: in the first case the property of the phenomenon corresponding to s_1 "implies" the property corresponding to s_2 (at the same time), in the second case the property corresponding to s_1 "causes" the property corresponding to s_2 (at a later time)). The correctness of this prediction is then tested by observing or detecting whether the state of the actual system at the given time corresponds to s_2. If it does, the prediction is "verified". If it does not, it is "falsified".

The representation as a whole cannot be verified in this manner since this would suppose that you test all possible predictions derived by using the representation and this of course is impossible. (On the other hand, as Popper (1968) would argue, a representation is falsified as soon as only one of its predictions is falsified. In practice, however, this criterion is never applied: a representation is accepted if most of its tested predictions are verified; a few exceptions can always be tolerated).

The difficulty with this procedure of operationalization is the same as that with formalization: the operational definition of a property or representation state is never complete, or unambiguous. A part of intuition remains in the decision whether a particular operation is, or is not, a good characterization of a representation state.

For example, in psychology we might have a representation which predicts that children who are raised in a stimulating environment will become more intelligent than children raised in a less stimulating environment. In order to test

this prediction, we must first operationalize the terms "intelligence" and "stimulating environment". Intelligence is usually defined operationally by the scores attained on an IQ test, which consists of a list of problems whose solution is supposed to require intelligence. This allows us to detect the IQ or "degree of intelligence" of an individual.

In order to test the prediction we must now prepare children in a state corresponding to more, respectively less stimulating environments. This might be done by supposing that Eskimo children who live in the polar region are less stimulated than Western children living in a big city. It now suffices to compare the IQ scores of Eskimo children with those of Western children of the same age.

However we may find that on the tests used the Eskimo children score better than the Western children. Does this imply that the hypothesis is to be rejected, or that the representation is falsified? Not necessarily: it may be that the intelligence concept was defined too narrowly, and that by using different IQ tests, the scores of the Western children would be better than those of the Eskimo children; it could also be that the preparation was not sufficiently definite, that the differences between Eskimo children and Western children other than those related to their degree of stimulation influence the process (e.g. the Eskimo race might be intrinsically superior in intelligence to the Caucasian race, or the Eskimo education might be superior to the Western education).

In the latter case the experiment should be redesigned for different populations, e.g. Western children from big cities compared to Western children from the countryside. But this would not eliminate all possible factors influencing the difference in IQ test performance: e.g. children from the countryside might be healthier because of less pollution.

We may conclude that although the scientist will always try to make the relation between his representation and the external world as explicit as possible through the use of operational definitions and tests, this operationalization is in the best case only partial. The degree of operational determination will in general be smaller if the domain to be represented is complex and changeful, and if the degree of formalization of the representation (its well-structured or paradigmatic character) is small. We must keep in mind that in such ill-structured domains it is often dangerous to formalize or to operationalize too rapidly: it is much more difficult to change a formalized and/or operationalized representation.

If the knowledge to be represented as a scientific model is still in a state of flux, it is better not to make this model very rigorous, since this would reduce the possibility to integrate new insights, or to change the concepts and principles. In such a case the function of formalization and operationalization is primarily that of directing attention to the insights that are still missing from the representation. Such a formalization or operationalization is therefore only provisional.

For example, from the experiments about the correlation between intelligence and stimulation, we might conclude that our knowledge about intelligence and its relation with the environment is still insufficient for

constructing an explicit representation of it. Yet in order to reach this conclusion we had to make a provisional operationalization of intelligence through the use of IQ tests. So, in a certain sense, the only result of this operationalization was to show that it is too early to operationalize, but on the other hand the analysis of the experiment might provide valuable clues about how to correct this situation.

3.4 The problem of transdisciplinarity

A major problem facing present-day science is its lack of integration: it appears as though scientific knowledge is fragmented into ever more disciplines and subdisciplines, which are ever more difficult to co-ordinate. The historical origin of the distinction of disciplines, is a distinction of problem domains.

The "*problem domain*" of a scientific representation corresponds to what we have called the "environment" of an adaptive system. This environment might be defined as that part of the universe with which the system is interacting. This means that for the system, the rest of the universe can be ignored, i.e. it does not have to be represented, since it has no direct implications for the eventual survival of the system. (The concept of environment defined in this way is similar to the concept of "Umwelt" which is used for biological systems).

For example, the nitrogen in the air an animal is breathing, does not belong to the animals environment, since it does not interact with its physiological system. Hence, it is neither perceived nor acted upon by the animal, and there is no need to have any knowledge about it. However, we may state that in principle everything in the universe is interacting with everything else, albeit in a very indirect way. For example, the nitrogen in the air is assimilated by certain plants, and these plants are eaten by certain animals, so that the nitrogen is indirectly transferred to the biochemical substratum of the animals body. Yet the animal is completely unaware of this process.

We may conclude that the distinction between the parts of the world which belong to the environment, and those which do not, is not objective or absolute, but dependent upon the representation characterizing the system, which selects those features of the world which appear directly relevant to the systems survival, and hence are worth to be perceived and represented, and ignores the rest.

The same is true for scientific representations, i.e. paradigms or disciplines, which single out a particular part of the universe, i.e. their problem domain, and consider it as important, while neglecting the rest. The community of scientist working within a specific paradigm can be considered as a system whose survival depends on the solution of the problems lying within the domain. However this system is not distinct and independent as a biological system: its boundaries are not well-defined, and different scientific communities are generally overlapping. Therefore it is better to consider it as a part of a larger system, which is the community of all scientists, or even better, society at large.

Whereas the problem domain of the smaller community centered around a paradigm may have a well-structured representation, the problem domain of the larger system of society has a rather ill-structured representation, made up of an incoherent and often inconsistent collection of subrepresentations. This is due to what Kuhn (1962) calls the "incommensurability" of paradigms. This means that different paradigmatic representations are generally incomparable, and untranslatable, so that it is impossible to establish definite relations between them, or to arrange them in an integrated whole.

This can be easily understood by remarking that one of the basic functions of an intersubjective representation is to allow *communication* between the individuals who share this representation. So, each social representation functions as a communication code or language, by providing intersubjective symbols, and rules for the combination (syntax), interpretation (semantics) and application (pragmatics) of those symbols. Since different representations use different symbols or rules, there is no direct communication possible between individuals who use different representations.

For example, the words "time" and "mass" are subjected to different rules in the Newtonian and Einsteinian representations of physics, and thus have a different meaning, even though their symbolic form is the same. Two scientists who know only the Newtonian, respectively the Einsteinian paradigm, will have great difficulty to communicate about the physical phenomena denoted by the symbols "time" and "mass".

In practice, however, different individuals dispose each of a set of representations, some of which are completely different, whereas others overlap largely. Therefore, there will always be a communication at a certain level of intersubjectivity (i.e. shared representation), although only a part of the intended meaning will pass through the communication code. For example, the Newtonian and Einsteinian scientists may communicate about physical phenomena on the level of the English language instead of using a shared scientific representation, but it is clear that this communication will be much more ambiguous and confused, since the meaning of the words "time" and "mass" is rather ill-defined in ordinary English, compared to the languages of classical mechanics or relativity theory.

This difficulty of communication across the boundaries of paradigms or disciplines explains the lack of integration of the scientific world view: indeed, how is it possible to coordinate different views of the world if there is no unambiguous common language in which those views can be expressed and compared?

This difficulty is more than purely theoretical: in general the tackling of a complex, social problem (e.g. famines) requires the collaboration of specialists from different fields (e.g. agronomy, meteorology, sociology, anthropology, …). Due to the lack of an integrating framework, however, such multidisciplinary approach will generally produce a mere aggregate of partial solutions which cannot be applied to the problem in a coordinated way.

These problems explain why presently such a sharp need for *transdisciplinarity* is felt. In practice, however, only lip service is paid to the advancement of the transdisciplinary ideal: it is much more difficult to really integrate different scientific approaches than to talk about the need to do so.

Moreover, the issue is confused because of a misconception about scientific representations, namely the idea that such a representation would provide an isomorphic image of a part of the world. This implies that different representations of the same problem domain be mutually isomorphic, and hence should be equivalent. On the other hand it should be possible that representations of different problem domains be stuck together, like the pieces of a puzzle.

The only problem seems to be that you need a universal language which is able to express all aspects of the world. All the equivalent and non-equivalent representations together could then be expressed in this language, and the resulting description would provide a unified scientific world view. This philosophy was at base of the movement for the unity of science, which was inspired by logical positivism. It has become clear, however, that this program has failed. The reason for this failure is that there is no fundamental isomorphism between representation and environment, and hence there is no universal method for synthesizing two different representations of the same environment.

3.5 The need for a metarepresentation

We have emphasized several times that the relation between a representation and its environment is not one of structural similarity or isomorphism, but one of adaptation, that is to say of correspondence between the feedforward processes guided by the stable representation structure and the actual changes in the environment. The same goes for scientific representations. Therefore, the way to integrate different representations is not to analyze their isomorphisms with the outside world, but to analyze the mechanism through which a stable representation structure can be adapted to a changing environment.

If this mechanism is sufficiently understood, we shall be able to compare different representations, to see in what aspects they differ, to see what is still missing in their organization to make them adapted to a particular domain, and thus to organize, coordinate and construct different representations in order to apply them to a given domain.

This problem is sometimes called the *representation problem*:

how to find the representation that is best suited for tackling a given problem domain?

It is also related to the problem of *representation changes* (Korf, 1980):

how to transform a given representation so that it is better suited for a given problem domain?

A solution to this problem would require the construction of a *metarepresentation*, i.e. a representation of all possible adaptive representations, and the way they work. This metarepresentation would express what all different representations have in common, and would thus provide a framework for their integration.

According to this philosophy, if you want to achieve transdisciplinarity, you must not look for a universal language for describing the world as it, but rather for a metalanguage for expressing, comparing and redesigning the existing partial descriptions. The basic argument for this idea is the fact that there does not seem to be a representation structure which can be efficiently applied to all problem domains. This conclusion was reached independently by different approaches, such as artificial intelligence, and philosophy of science.

We come now to the point where we can reformulate the basic research problem with which this work is concerned:

how can we achieve a universal metarepresentation, that is to say, a representation of all the possible ways through which a changing environment can be represented so as to allow for adaptation?

Although our analysis of the representation concept has contributed much to a better formulation of this problem, the problem domain remains rather ill-structured. Translated in the language of philosophy of science, this means that the study of this problem is still in a pre-paradigmatic stage. This implies that there are as yet no explicit rules or evaluation criteria to guide the search for a solution. In particular, there is not any general formalization or operationalization available which could be applied to the "theory of representations" (cf. Simon, 1979) we are looking for.

Until now the most explicit approaches to this problem can be found in Artificial Intelligence (Korf, 1980; Amarel, 1968), but as we have shown the problem transcends computer science proper by its transdisciplinarity. From my point of view, what is missing in the existing AI approaches is the emphasis on change and adaptation. It is here that the physical sciences, in particular theoretical physics and systems theory, come in. What we need is a synthesis of the cognitive approach (internal structures and processes of a representation) and the physical approach (adaptation to external change). Therefore we must first analyze as profoundly as possible the existing theories and concepts of these two approaches.

This will be carried out in two stages: first, we will analyze what will be called the "classical representation frame", second, we will criticize this classical frame by reviewing a number of more recent attempts to go beyond it.

The main difficulty here is that this review of the literature must encompass a very large range of disciplines and models, so that it can at best be very incomplete. In particular, since it is impossible to be an expert in all these different domains, it cannot be guaranteed that the topics chosen from those different fields are the most important, or that their description is the most accurate. The best one can do is to make a subjective selection of those aspects which seem to fit best to the general problem, and to hope that the way they are rendered will not be too much criticized by the experts.

In parallel with this analysis and overview of existing ideas, we will begin to construct a new metarepresentational framework, which is more general than the classical framework, while integrating the various new approaches which were reviewed. As was argued, in this pre-paradigmatic stage of research, it is too early to attempt a complete formalization of the theory. Yet, where possible, we will indicate how the ideas may be expressed in a formal way, so that the number of misconceptions due to vagueness or ambiguity can be restricted as much as possible.

An even more difficult problem is operationalization. It is clear that the traditional procedures of operational testing (quantitative agreement of theoretical and experimental results) are inadequate for evaluating a metarepresentational theory. Yet the theory is supposed to be applicable to concrete situations. Therefore we must look for alternative means of testing the adequacy of the theory. A possible approach to operationalize the theory will be sketched in sect. 11.7.

The basic thrust of the present approach, will, however, remain theoretical. This is in accordance with Kuhn's (1962) observation that pre-paradigmatic or revolutionary science is characterized by a philosophical questioning of the very foundations of the knowledge involved, and by an attempt to synthesize the available data in a new, integrative framework.

CHAPTER 4: A reconstruction of the classical representation structure

4.1 Introduction

If we wish to study the fundamental structures of adaptive representations, it is best to begin with those representations which are most explicit and most generally used. These representations can be found in what I would call "*classical*" science. The prototype of such form of science is the theory of classical mechanics. I hope to show, however, that the same basic structures which can be found in classical mechanics are also used in very different scientific representations, such as logic, computer science and linguistics.

Science was defined as an attempt to construct explicit representations. However there remain many presuppositions in scientific theories which are implicit. Usually these have the form of principles which are so evident that no one makes the effort to think about them, to question them, or even to formulate them. But science can only progress if things which were first intuitively evident, are later carefully expressed and examined, so that they can be integrated in a more general predictive model.

The traditional illustration of this principle is the story of the discovery of gravitation by Newton. Before Newton, everybody knew intuitively that heavy things tend to fall to the ground, and there seemed to be no reason to search for an explanation for this fact. Newton's basic insight, however, was that there must be something, a force, which attracted heavy bodies towards the center of the earth. This concept of gravitational force then allowed him to construct a theory which not only explained falling bodies, but also the movement of the planets around the sun. This was the first step in the formulation of the paradigm of classical mechanics.

What should be done here, however, is to go one step further, and to formulate and examine explicitly those intuitive principles which determine the structure of the classical representation, and so to uncover its hidden constraints and limitations. This would be the first step in the construction of a new framework which would be more explicit, more general and more flexible, and hence applicable to a much larger range of problems.

This analysis will be undertaken as a reconstruction of the more advanced representation structures from their most simple elements. So, the study will follow an approximately historical or developmental sequence, proceeding from the most primitive to the more elaborated cognitive structures. This approach is somewhat similar to the "genetic epistemology" program initiated by Piaget (1972), and which was applied to the development of artistic representations by Blatt (1984). The structural analysis of classical science will make use of certain formalisms used by logical empiricists (e.g. Carnap, 1958) and by researchers in the field of problem solving (e.g. Nilsson, 1971; Amarel, 1968).

After analyzing these structures, we will try to demonstrate what the use of these particular cognitive structures implies for the domain of applicability of the classical representation. In other words: what view of the world is associated with these classical representation structures (see chapter 5). The purpose is to show that if you wish to change your world view or metaphysics (i.e. the content of your representation), then you must also change the form of your representation.

We will now begin the reconstruction of the classical frame by studying its most basic level, that of its primitive elements: objects and predicates.

4.2 The generation of elementary expressions

The roots of the scientific representation frame lie in the representations provided by verbal language. The mechanism of verbal language is typically human: no other species of animals disposes of a system of signs which is at once *discrete* and *infinite*. Signs used for animal communication are either finite in number (e.g. a finite set of warning signals for different situations of danger) or continuous (e.g. the continuously varying speed of the movement of a bee can indicate the variable distance of a source of food).

This feature of discrete infinity results from a finite set of *generating elements* (words) which may be combined in unlimited sequences by *generative rules* (grammar) (Chomsky, 1972). The mechanism presupposes that the generating elements are stable, that they do not change in form or content during the operation of combination (i.e. sentence formation). These elements are units of representation or meaning: each element has a given, invariant meaning which cannot be analyzed further.

The combination mechanism, governed by grammatical rules, is basically reducible to a *linear concatenation* of elements. (The linearity of verbal expressions can be understood by the limitations of human speech: it is impossible to utter several sounds at once, so the different signals must succeed each other in a sequential fashion).

The most simple linear concatenation consists of an ordered pair of elements:

a.b, or a (b).

This is a model of an atomic proposition or sentence: something (a) is said about something else (b).

Concatenations consisting of only one element (e.g. b), are not really assertions; they only direct attention to the phenomenon denoted by b, but they do not relate it to other phenomena. Therefore, a language consisting of only unary expressions (e.g. a, b, d, …) is unable to represent change: it is impossible to describe any (dynamical) relationship between the different expressions of the language. Such a language would lack the coherence and continuity required for the organization of adaptive representations.

The simple binary coupling of two elements a(b) is sometimes called "*predication*": a is predicated over the subject b. Remark the basic asymmetry of the relation between a and b: a(b) is different from b(a). Therefore, a and b have different functions in the expression. The function of a can be called "*predicate*", "attribute" or "property", that of b can be called "subject", "individual", or "*object*". In order to visualize this difference we will use capital letters for the predicates. Therefore, we get: A(b). The simplest interpretation of this expression is: object b has the property A.

In a more complex type of proposition, the subject of the predication consists of several elements: e.g. A(b, c). In that case the predicate A is interpreted as a relation between the objects b and c: e.g. if b and c represent numbers, then A(b, c) may signify: "b is smaller than c".

These predicative propositions form the basis of the formal languages used in classical science (see e.g. Carnap, 1958). They also form the skeleton for sentence formation in natural language. This may be illustrated by the examples of two prototypical sentences:

1) "John is a farmer" can be reconstructed as: **Farmer (John)** (monadic predication)

2) "John beats the dog" has the form: **Beats (John, dog)** (dyadic predication, or relation).

There is also reason to believe that this predication scheme forms the basis of conceptual thought and representation. As such it is used in the existing AI models of knowledge representation (see e.g. Charniak and McDermott, 1985). We will now try to analyze how this scheme can be used in the representation of change, and what are its qualities and limitations. First we must analyze the functions of objects and predicates.

4.3 The function of objects and predicates

As we have said, to be useful the elements of the representation we are studying must be stable with respect to form and content. The form is determined by the words or the symbols of the language we are using. The content or meaning of a representation has two dimensions: the *extension* (i.e. the external phenomena that are denoted) and the *intension* (i.e. the concept or interpretation of the phenomenon). A representation is defined as a relation between intension (self) and extension (world) (sect. 2.1). Therefore, an elementary expression should connect an extension with an intension.

Since a separate element is stable, change must be represented by the *variable combination of invariant elements*. But if we wish to adapt to change, that is to say make anticipations, we must suppose that there is always something

continuous in the process, something which establishes a relationship between the present and the future. Therefore, the transition from the present state of a representation to the next one should conserve certain features of the state.

The simplest way to do this in the object-predicate scheme is by varying the predicate, while keeping the object constant. This corresponds to the basic principle that in the environment there are certain invariant "entities", which can be represented as objects. The changes in the environment can then be reduced to changes in predicates, i.e. transformations of the "properties" or "relations" of those entities.

The supposed invariant external entities constitute the stable extension of the representational objects. The predicates on the contrary have no stable extension since the phenomena they are supposed to represent (i.e. properties such as velocity, relative position, color, etc.) are variable. As a way of "compensation", the predicates have a more invariant intension than the objects: the same predicate can be attributed to different objects, while retaining the same internal meaning. In this sense they are "universalia" rather than "particularia" like the objects. Hence, we may postulate that an object expresses a stable extension, and a predicate a stable intension.

This allows us to reduce the distinction between objects and predicates to the subject-object dichotomy or the self-world duality. An object is indeed supposed to be completely determined by the external world. As such it is "proposed" to the self or subject who will try to describe it by attributing predicates to it. But these attributions presuppose a perceptual system able to interpret the signals coming from the object, and to organize them in distinct categories. Hence, the attribution of a predicate depends on the representation state and structure of the perceiving subject.

Let us illustrate these abstract considerations by means of a few examples. Let us begin with two different phenomena, a *billiard ball* and a *cloud of gas molecules*, and check whether they can be adequately represented as objects.

In the case of the ball there seems to be no problem: its physical structure is quite stable, and it can undergo different manipulations (e.g. translations, rotations, …) without loosing its recognizable identity. Hence, it is easy to describe different processes involving billiard balls (e.g. the different strokes occurring during a billiard game) by a sequence of propositions, which are constituted by an invariant object (the ball) and a variable predicate (e.g. the position, the velocity and the spin of the ball).

This is no longer true for the cloud of gas molecules. Unless the cloud is confined in a box, all processes involving the cloud will be accompanied by a natural diffusion of the gas molecules. This means that at any moment form and volume of the cloud are changing. The individual molecules will be scattered in all regions of space and will get mixed with other molecules of the same or of different substances. Within the shortest time interval it will become practically impossible to separate or to distinguish the original cloud from its background.

Clearly the extension of the "object" cloud is not stable: after a very short time it is impossible to check whether a particular phenomenon (e.g. an isolated gas molecule) belongs to the original extension denoted by the representational object "cloud". Therefore the representational element "cloud" cannot be used in an object-predicate scheme of representation: it is operationally meaningless to attribute predicates (e.g. volume, or mass) to the cloud, since there is no way to observe in what way these particular attributes change, without interference from other phenomena (e.g. other clouds which exchange molecules with the original cloud and in this way influence its mass and volume).

From this example you may conclude that if you wish to construct a (classical) representation of a certain environment, then you must be careful in distinguishing the phenomena occurring in that environment that will be represented as "objects".

The same thing can be said about the predicates. The basic requirement for a representational element to function as a predicate is to have a stable intension. This means that there must be an invariant perception and interpretation scheme, which categorizes different phenomena which are similar in a particular respect, in one and the same category. In particular, two phenomena which are identical in all respects relative to the representation should be attributed the same predicate.

For example, two cars which were manufactured following the same fabrication procedure should have the same length. This means that you or I when measuring our cars with a meter stick, should get the same results. Therefore, we may say that the cars can be attributed the property of being e.g. 4 meters long. There is no ambiguity about the observation and interpretation operation necessary for establishing this fact.

This is however no longer true for e.g. the predicate "Beautiful": whereas you may consider that the car is beautiful, I may think it is ugly. Clearly, the predicate "Beautiful" has a different intension for you and me. Moreover, whereas you may find the car beautiful when seeing it exposed in a nicely arranged showroom, you may find it ugly after you were almost knocked down by it in a dark street.

Apparently the intension of the representational element "Beautiful" is rather variable. Therefore it can only be used in restricted cases, and is not very well suited for general descriptions of processes.

What we have called the stability or invariance of extensions and intensions means in practice that different phenomena are mapped onto the same representation unit, and hence are considered equivalent or identical with respect to the representation: they are "*assimilated*" (see further sect. 6.2).

In the case of extensions it means that an external phenomenon is represented as an object with an invariant identity, even though the phenomenon passes through different states. From the viewpoint of the identity of the object, these states are considered equivalent.

In the case of intensions, it is an internal process of perception or interpretation that is represented as an invariant predicate, even though different phenomena which trigger this process will differ in certain aspects. Here too these aspects are neglected, so that the representations become equivalent. In general each representational state or unit represents an *equivalence class* of phenomena which are mapped onto it.

The fundamental problem here is to determine the *boundaries* of this class: when does a phenomenon belong to that class, and when does it no longer belong to it? For example, if a billiard ball breaks in two, must it still be represented by the same object "billiard ball"? If the rod functioning as a measuring unit is deformed, then can it still be used to attribute the predicate "Has a length of one meter"?

Clearly there are limits to the stability or invariance of representational elements, beyond which the element loses its adequacy for representation. In the classical representation frame there is an implicit presupposition that the phenomena which are represented, always remain within these limits, so that the same representation units (objects and predicates) can be used throughout the process of change that is represented.

We must now analyze the structure of the domain enclosed within those limits. Later on we will study what happens when the boundaries of these domains are transgressed.

4.4 The generation of compound expressions

We have seen that an elementary (or primitive) expression can be generated by the coupling of a predicate to one or more objects. In general however, the description of an outside situation requires a more complex representation. This can be achieved by combining different primitive expressions into a *compound* or "molecular" expression by means of *connectives*.

The fundamental connective is the operator of *conjunction*: it expresses the fact that two expressions are both true descriptions of the given situation. We will symbolize the conjunction operator by a point (.). For example:

Farmer (John). Beats (John, dog)

signifies that John is a farmer *and* that he is beating the dog.

The relation between two expressions connected by a conjunction is symmetrical: a commutation of the sub expressions does not change the meaning of the compound expression. The number of atomic expressions that may be connected by conjunction is unlimited. Hence, it can be used to construct arbitrarily complex descriptions of a particular situation by listing all of its relevant features (as far as they can be expressed within the predicative language).

The second basic operator for making compound sentences is *negation*. The underlying idea is that either an expression of the language is an adequate

description of a certain aspect of a situation, or it is not. There are no intermediate cases where a representation would be partially correct. This principle derives from the basic presupposition of invariance of the representational elements. This invariance defines a domain of equivalence, in which all phenomena can be represented by the same invariant representation, whereas the phenomena which fall outside this domain cannot. Hence, we get an absolute *distinction* or dichotomy between correct an incorrect, or between true and false, instead of a continuous sequence of degrees of adequacy.

This distinction can be used to associate any expression s with its negation s' or NOT s, which would represent all those situations for which s would be incorrect. Remark that this procedure is rather artificial since there is in general no positive way to characterize the set of situations represented by s'.

With the help of the connectives "." and "NOT" other connectives can be defined:

disjunction:
a OR b is defined as NOT (NOT a. NOT b);

implication:
a ⇒ b is defined as (NOT a) OR b;

equivalence:
a ⇔ b is defined as (a ⇒ b). (b ⇒ a).

These derived connectives can be used to simplify compound expressions, but they do not add anything fundamentally new to the representation. The properties and applications of the different connectives can be found in any book on formal logic (e.g. Carnap, 1958). What interests us here, are not so much the formal properties, but the cognitive implications of this representation structure.

4.5 From Boolean algebra to state space.

The set of objects (O), the set of predicates (P) and the connectives for constructing compound expressions together determine a set of possible expressions (E). This set is defined recursively:

an expression e belongs to E, if it can be constructed by the combination with the help of the connectives, of the expressions e_1, … e_n, with e_1, … e_n belonging to E.

the primitive expression A (b) formed by the combination of a predicate A, belonging to P, with an object b, belonging to O, is an element of E.

If the connectives used can be reduced to the conjunction and negation, as defined above, E has the structure of a *Boolean algebra* or *Boolean lattice* (Mac

Lane & Birkhoff, 1967; Halmos, 1974). This means that you can define a *preorder* < on E such that:

e < f *iff* (if and only if) e ⇒ f is always true, for every possible truth value of e or f.

Intuitively this means that you can derive or deduce f from e: if you know that e is a true description of a situation, then you also know without further observation that f is a true description. You could say that the information about the situation carried by f, is already contained in e, f does not add anything new to your information. For example, if you know that a. b is true, then you also know that b is true, hence a.b < b.

The operations of conjunction and disjunction function as greatest lower bound, respectively least upper bound for the partial order determined by the implication. Both operations are related by the property of *distributivity*. These algebraic properties are typical for the logical structure of the classical representation. They allow us to relate different expressions in a static way.

Within this framework we may now discuss the concept of "*state*". According to the general definition that was formulated earlier (sect. 2.5), any expression belonging to E can be a state of the representation, i.e. an actual combination of representation units selected from a larger set of potential combinations. As such it represents the information we have got about a particular outside situation. Therefore, the state characterizes the actual features of the subject's awareness of the object.

In the classical representation, however, one tends to think about the state concept as a characterization of the object itself, independently of the subject. (These "objective" features of the environment at a particular moment, are what we call a "situation"). This means that the limitations implied by a subjective viewpoint (incomplete information, biased perception of features) are ignored.

Therefore the *information* carried by a "classical" state is supposed to be *maximal*: in principle, the state should completely determine or characterize the object(s) belonging to the domain of study. This is exemplified in classical mechanics by the position of a system in phase space: if you know this position, then you can deduce all observable features (energy, velocity, momentum, …) of the system, and find a unique solution for the dynamical equation determining its evolution.

If we translate this requirement into the language of logic, we get that the expression corresponding to a classical state should be such that the truth or falsity of all the other expressions can be derived from it. In a Boolean lattice, this corresponds to the property of "atomicity" of the state expression (Piron, 1976):

the expression s is an *atom*, *iff* (if and only if) 0 < s implies that 0 is a minimal element of the lattice (this means that the expression 0 is never true).

It has the property that for any expression a we have:

either s < a, or s < NOT a.

(This property is no longer valid for atoms in a non Boolean lattice, see sect. 7.2).

The < relation can be interpreted as an ordering of expressions according to the amount of information they provide. The states or atomic expressions are then the expressions incorporating the *maximal amount of information*. As such they are mutually incomparable; there is no logical or informational order relating them. This leaves us with an unstructured set S of state expressions, which will be called the *state space*.

The remaining expressions can be reduced to combinations (classically disjunctions) of state expressions or atoms. As such any expression e from E corresponds to a subset S_e of S (cf. Piron, 1976):

$S_e = s_i \in S$ such that $s_i < e$.

Thus we see that the introduction of a state as a maximally determined expression in a Boolean lattice allows us to reduce the lattice of expressions to a state set, and its subsets. The remaining problem of how to represent change can now be restricted to the introduction of a dynamical structure on the state space.

Let us first illustrate these concepts with the help of the billiard example. The representation of the billiard situation starts with the selection of one (or more) billiard balls lying on the table as basic objects. As basic predicates we can distinguish different positions for the ball. The expression s: P(a) means "the ball a is in that part of the table that is labeled P". A compound expression t might be P(a). Q'(a). NOT R(b), i.e. "the ball a is in the intersection of the regions P and Q, whereas the ball b is outside the region R". Clearly we have the implication t < s.

The atoms or states of the billiard game representation are those expressions where all the balls are localized with maximal precision on the table. This precision depends of course on how fine-grained the system of predicates (or regions) is. In the ideal case, where the precision is maximal, each ball would be localized at a particular point. The set of all points within the boundaries of the billiard table would then determine the state space for one billiard ball. The state space for several billiard balls may in this case be obtained by taking the Cartesian product of the individual state spaces. An enlarged set of predicates, and hence of states, could be obtained by including the momentum or velocity as descriptive features of the balls. In that case the resulting state space would be equivalent to the phase space of classical mechanics for the system of billiard balls.

4.6 Topology, time and trajectories in state space.

The interesting feature of a classical state space is that you have a set of descriptions of possible situations which are mutually exclusive: for s, t ∈ S, s ≠ t, you have that either s or t is an adequate description, but not both of them; s.t is never true.

If s and t would be completely written out as conjunctions of primitive expressions and their negations, we would normally find that one of the positive sub expressions in s, would occur in t in negated form (this is a general property of Boolean algebras). Therefore, to go from s to t we must at least negate one proposition which was true for s, but is no longer true for t.

This replacement of a primitive expression describing part of a situation, by its negation can be conceived as the basic mechanism for representing change within the classical frame. Therefore, any process of change is basically decomposed into a sequence of elementary, discontinuous transitions, in which a primitive proposition is exchanged for its negation.

Yet a fundamental property of change as we experience it, is its generally continuous nature. This *continuity* can basically be formulated by saying that for time intervals which are not too large, the smaller the interval, the smaller the change occurring within that interval. In order to state this more explicitly we must define what is meant by "small changes". A small change could be said to be a change which does not take its object "too far away" from its original characterization, so that its new state is still in the "*neighborhood*" of the previous state.

This concept of neighborhood is the fundament of *topology*, which is a mathematical framework for the study of continuity. Therefore, if we want to describe change more realistically, we should begin by providing the state space with a topological structure. This can be defined alternatively by a family of "neighborhood" sets, a family of "open" sets or a family of "closed" sets. A function is then defined to be continuous if it conserves this topological structure.

In practice, however, the "continuity structure" of a classical state space is not logically build up from the foundations provided by topology. The possible structures characterizing such a space are complex and diverse: topological structures, projective structures, linear structures, metrical structures, …

The historical origin of these different spatial structures can be traced back to the Euclidean model of 3-dimensional geometry. This structure is very rich and elaborate, and could be viewed as the structure of the state space of a "material point" or "atom", which is an extreme idealization of the concept of an "object". The historical evolution of geometry, starting from the Euclidean model, has gradually analyzed, generalized and decomposed this model into ever more primitive substructures, which were elaborated independently. The result is that the present state of geometry is rather complex and confuse, and incorporates notions from algebra, calculus, topology, and so forth.

One of the issues leading to confusion is that the concept of space is as well applied to the network of relations between actual objects, as to the network of relations between potential states of one or more objects. The latter concept is what we call a state space, the former corresponds to our intuitive notion of space, as experienced through visual and bodily perception. We will focus here on those aspects which are directly connected to dynamics.

The classical way of viewing change is to reduce it to "movement", i.e. the continuous evolution of the position in (state) space of the observed system. The structure of the space is then determined by the structure of the set of possible "movements". Traditionally, however, the geometrical structure of space is supposed to be given beforehand, i.e. before change is introduced. In order to make the transition from the static representation of space, to the dynamic representation of movement or change, the classical frame needs one more representational structure: *time*.

Classical time is basically a *linear order relation* between different actual states. This relation will be denoted by P:

s_i P s_j or P (s_i, s_j) (s_i, $s_j \in S$) reads: "the state s_i precedes the state s_j".

P has the relational properties of antireflexivity, antisymmetry, transitivity and completeness. (completeness: for all s_i, $s_j \in S$: s_i P s_j or s_j P s_i).

This means that the different states of the changing system you wish to represent can be ordered in a *linear sequence*. This sequence can be indexed by labeling the consecutive elements of the sequence with numbers (integer or real) from a number system. These numbers correspond to the chronological time you read from a clock when the particular state that is labeled, takes place.

In order to describe the evolution of the state it now suffices to give a set of states as a function of time:

$s_1(t_1)$, $s_2(t_2)$, $s_3(t_3)$, … with $s_i \in S$, $t_i \in T$,

where T is a number system (integers or reals) symbolizing time. The constraint is that:

$s(t_i)$ P $s(t_j)$ *iff* the number t_i is smaller than the number t_j.

The parameterized sequence s(t), $t \in T$ is called the trajectory of the system in state space.

The order relation on the set T of moments determines a topological structure; the neighborhoods of this topology are the open intervals $]t_1, t_2[$:

$\forall\, t \in T, t_1 < t_2 \in T: t \in]t_1, t_2[$ *iff* $t_1 < t$ and $t < t_2$

This topology can be "coupled" to the topology of the state space by demanding that the function from T to S, which maps a time $t \in T$ onto the state s(t) at that particular time, be *continuous*. This means that for any neighborhood $O(s_0)$ of a state $s_0 = s(t_0)$ in S, there must be a neighborhood $O(t_0)$ of $t_0 \in T$, such that if $t_1 \in O(t_0)$ then $s(t_1) \in O(s_0)$.

Intuitively, it means that for a time t_1 "close to" t_0, the corresponding state $s(t_1)$ must be "close to" $s(t_0)$; if $s(t_1)$ would be "far away" from $s(t_0)$ then we could find a "smaller" neighborhood $O'(s_0)$ such that $s(t_1)$ does not belong to it, even for t_1 arbitrarily "close" to t_0 (i.e. for $O(t_0)$ arbitrarily "small").

This requirement of continuity of trajectories determines a direct coupling between the temporal order relation and the topological structure of state space. It signifies basically that before a large change can take place (i.e. the trajectory crosses the boundary of a large neighborhood of the initial state $s(t_0)$ at time t_2), a smaller change must have occurred (i.e. the trajectory must have crossed the boundary of a smaller neighborhood of $s(t_0)$ at a time t_1, with $t_0 < t_1 < t_2$).

Once again we will illustrate these concepts with the example of the billiard game. The states $s \in S$ of the billiard game are the Cartesian products of the positions of the individual balls on the billiard table. The evolution of the game can now be modelled by a parameterized sequence s(t), $t \in T$, of states. Each time-dependent state s(t) represents the position of the different balls at the time t. Remark that this presupposes that the exact positions of two separate balls can be established in one and the same instant, i.e. the parameter t is the same for all the different balls. This is called "absolute time".

The continuity of movement requires that the balls cannot jump over finite distances in an infinitely small time interval. In order to express this you need a topology on the set S of states. This is traditionally derived from a metric or distance function on S:

$d: S \times S \rightarrow R: (s_1, s_2) \rightarrow d(s_1, s_2)$,

which is itself constructed from the metrics on the component spaces of S which represent the possible positions of the individual balls. These component metrics are determined by the postulated Euclidean geometry of the component spaces.

The neighborhoods needed for defining the topology are the open balls in state space:

$B_r(s_0) = s \in S: d(s, s_0) < r, r \in R_0$.

Continuity of evolution is now expressed by demanding that for any distance r however small it may be, you can find a time interval t_1-t_0 sufficiently small such that:

$d(s(t_1), s(t_0)) < r$.

In other words, how shorter the time the billiard balls are observed to move, how smaller the distance they will have traveled.

4.7 The group of dynamical operators

We have seen how a given evolution of a system can be represented as a state space trajectory parameterized by time. We now also wish to represent the way such evolution can be brought about or manipulated. The actions causing a change of state will be called operations. Their formal representations in the state space framework will be called *transformations* or *operators*. These are basically functions mapping a part D of the state space S onto another part C. In the classical frame the functions $f \in F$ are one-to-one: C = D = S. The basic idea is that a state s is transformed by an operator f in a state s':

$$f: S \rightarrow S: s \rightarrow s' = f(s)$$

f can represent an action, force or influence exerted by the observer or the environment upon the system, causing it to change in a specified way.

The interesting feature of these operators is that they can be composed: the composition of two operators f, g produces a new operator $h = f * g$ which represents the action consisting of carrying out the action represented by f after the action represented by g. The composition operation introduces an algebraic structure into the set of operators. Classically this structure is that of a "*group*". This means that you have four special properties:

1) composition is *everywhere defined* in the set F of operators:

$$\forall f, g \in F, \exists h \in F: h = f * g.$$

This signifies that the carrying out of an action or operation is independent of the history of the system, i.e. of the previous operations carried out on the system. A given operator f is applicable to any state s_1 of the system, whatever the way it was brought about. The resulting new state s_2 can be subjected to any other operator g, leading to a third state s_3. The succession of the two operators f and g can then be represented by another operator h, which leads directly from s_1 to s_3.

2) composition is *associative*:

$$\forall f, g, h \in F: f * (g * h) = (f * g) * h = f * g * h$$

This property signifies basically that if more than two operators are to be composed, then composition can be carried out sequentially: first you compose g with h, then you compose f with the result of that composition; it does not

make a difference if f is first composed with g and their composition is then coupled to h. This sequentiality corresponds to the linear order of time.

3) there is an *identity* element $i \in F$ such that:

$$i * f = f * i = f, \forall f \in F.$$

This property corresponds to the idea that one possible action is to do nothing at all. The operator i representing this action maps every state onto itself:
$\forall s \in S: i(s) = s$.

Carrying out an operation f and then doing nothing at all, is equivalent to only carrying out f.

4) every operator $f \in F$ has an *inverse* $f' \in F$ such that:

$$f * f' = f' * f = i$$

This property presupposes that every action can be undone. Carrying out f and then carrying out its inverse f' is equivalent to doing nothing at all. All operators are reversible, nothing has been irrevocably lost or damaged, you can always get back to the situation you started from.

The group F containing all operators defines an *equivalence* relation on the state space S:

"s_1 is equivalent to s_2" *iff* there is an $f \in F$ with $f(s_1) = s_2$

This relation is reflexive, due to the identity property, and symmetric, due to the reversibility property. It is transitive because two operators can always be composed. Therefore, it is an equivalence relation which partitions S into disjoint equivalence classes.

A further presupposition of the classical representation frame is that if you consider F as the group representing all physically possible operations, then there is only one equivalence class, which is S itself. This property signifies that any state can be transformed into any other state of S by applying the right operator. Therefore, it suffices to start with only one state s_0 in order to construct the whole state space by applying all operators $f \in F$ to this one state.

You can say that the group F defines the state space of the system, and hence the system itself. Carrying out an action which is not represented by an operator $f \in F$ would mean that the state would be sent upon something "outside" the state space, i.e. the resulting situation would no longer be represented as a state of the system, but possibly as a state of another system. So, the only "invariant" of the transformation group F is the system itself.

Although the operators belonging to F represent possible changes of the systems state, these changes are clearly not the most general. The changes

represented by the group F may be called “changes of the first kind” (cf. Watzlawick, Weakland and Fisch, 1975), i.e. changes which leave the system itself invariant (the system is here defined as the set of objects, whose properties and relations determine the space of possible states). “Changes of the second kind”, i.e. changes of the components or structures of the system, are not represented by elements of the group.

This feature of changes “remaining within the system”, in contrast to changes “which jump out of the system”, is a consequence of the property of “closure” which characterizes groups: anything happening in the group remains within the same group, it can never “leave” or “get out of control”. Indeed, any possible combination of operators of the group, representing a complex of actions performed on the system, can be reduced to a linear sequence of elements by associativity. This sequence can be simplified by composing its first element with the second element, then taking the resulting operator and composing it with the third element, and so forth, until we are left with just one operator, representing the whole complex of actions. Because of the reversibility property we can find an inverse of this operator. This means that whatever complex combination of actions we perform on the system, there is always one operator which is sufficient to bring the system back to its initial state.

Hence, any state produced by any combination of operations from an initial state, can always be directly retransformed in this initial state. The state space and the corresponding group of state space transformations, are closed; they form one separate, indivisible whole, an equivalence class.

Let us now go back to the billiard game example. If our representation of the game consists of a set of N balls (objects), characterized by their positions on a two dimensional, infinitely extended billiard table (predicates), the possible actions in the game will be represented by changes of position of the balls, i.e. translations in the two dimensional space of individual positions, corresponding to translations in the 2N-dimensional space of states of the system of balls.

These translations form a group: 1) the composition of two translations is another translation, 2) the composition of three translations in a row is the same, whether we begin by composing the two first translations and then adding the third one, or by composing the two last translations and then composing the result with the first one, 3) there is an identity translation, which does not move the balls at all, 4) for any translation there is an inverse translation, which moves the balls over the same distance but in the inverse direction.

Remark that if we had a regular billiard table with a finite length and breadth, the possible translations of the balls on the table would not form a group. Indeed, if we would compose ten translations which would each move a ball by 50 cm in a specific direction, the resulting translation would move the ball by 5 m, and hence transgress the tables boundary. In this case the ball would literally “jump out of the system”.

If we nevertheless wish to define a transformation group for a finite billiard table, we must introduce operators which are no regular translations. (These operators may for example be characterized by requiring that when the

trajectory of a ball which is being translated reaches the edge of the table, its direction should be reflected in such a way that the reflected trajectory forms an angle with the edge equal to 180° minus the angle of the original trajectory).

As we have already discussed, this “kinematical” representation of the billiard game way be extended to a “dynamical” representation by including the velocities or momentums of the balls as supplementary predicates. We then have to define operators acting upon these new components of the state. These operators can be thought of as accelerations or forces, which change the velocities of the balls. The actions represented by these operators could for example be the different strokes exerted on the balls with the help of a billiard-cue.

The set of accelerations has again a group structure: the composition of two accelerations is an acceleration, composition, is associative, there is an identity acceleration, and each acceleration can be inverted.

The transformation group of the billiard game as a whole, is given by the Cartesian product of the translation (possible with reflection) groups of the individual balls and the acceleration groups of the individual balls. This group is sufficient to represent all possible changes of state occurring in a regular billiard game (if we ignore for the moment the spin properties of the balls).

However, we can easily imagine changes of the billiard system which, although improbable, are physically possible, but cannot be represented by operators of this group. For example, one of the balls might break in two. This would change the number of objects needed for the representation of the system, and thus would add another two (or four) dimensions to the state space by increasing the number of degrees of freedom. In a certain sense the billiard game would “jump out of its state space”. There is no way to recuperate this change by extending the group, since the process of breaking in two is clearly not reversible, hence cannot be imbedded in a transformation group.

Another example of such a change of the second kind would be the introduction of an obstacle on the table-cloth. This would change the geometry of the state space and hence of its corresponding invariance group. An even more drastic change would be the melting of the balls. This would change the representation from one corresponding to the mechanics of material points or rigid bodies, to one corresponding to continuum mechanics, characterized by an infinite number of degrees of freedom.

4.8 Dynamical constraints

We have analyzed the basic mechanisms used in the classical frame for the representation of possible changes occurring in a system. For a representation to be adaptive, however, it should allow us to make predictions, i.e. it should not only model possible changes, but also the actual changes which are to be expected. In order to do this, we need some mechanism for restraining the number of changes to be expected in a given situation. Such mechanisms may be called dynamical constraints or laws. In the ideal case they allow us to select

a unique trajectory, i.e. a time-parameterized sequence of operators. In general, from a given initial state $s(t_0)$ several subsequent states s_i (t_0+T) can be reached by applying different operators $f_i \in F$:

$$f_i\,(s_0) = s_i.$$

The only constraint given as yet is that of continuity: for a small interval T, the subsequent states s_i $(t_0 + T)$ should be in a small neighborhood of s_0 (t_0). This already restricts the number of operators applicable for representing the changes in a given time interval T. In order to make useful predictions, however, we need a further reduction of the set of applicable operators.

There are two basic types of constraints for selecting the right operators: *conservation* (or invariance) *principles*, and *variation* (or optimization) *principles*.

Those of the first kind express the requirement that a given property of the system should be conserved during the system's evolution. An example of such a principle is the conservation of energy in classical mechanics. Affirming that a system has a certain property, which is invariant, means that there is a certain proposition, consisting of a combination of objects, predicates and connectives, which is implied by all states the system can reach during its evolution.

This proposition e_0 (which is for example equivalent to "the system's energy is equal to E_0 ") will in general be true for a certain set $S_0 \subset S$ of states, and false for all the other states. The conservation principle demands that only states belonging to S_0 can lie on the system's trajectory. Therefore, this constraint amounts to a reduction of the state space.

Since all states of the original space S can be reached from a given initial state $s_0 \in S_0$ by applying the appropriate operator $f \in F$, this implies that also the set of operators should be reduced to $F_0 \subset F$, such that:

$$\forall\ f \in F_0\colon f(s_0) \in S_0.$$

It is easily shown that F_0 is again a group, hence a subgroup of F:

1) composition is internal in F_0: f, $g \in F_0$ means that f and g conserve the property e_0. Clearly then $f * g$ will also conserve e_0, hence $f * g \in F_0$.

2) composition is associative in F_0 since it is in F.

3) the identity element i conserves all properties, hence it conserves e_0 and belongs to F_0.

4) if f conserves e_0 then its inverse f' will also conserve e_0, hence $\forall f \in F_0$, $f' \in F_0$.

We may conclude that, in a certain sense, the introduction of a conservation principle simply reduces the representation to a smaller representation, determined by a state space S_0 which is a subset of S, and a transformation group F_0 which is a subgroup of F. This may seem rather trivial. You might be tempted to ask: why bother about the conservation principle? Why could we not simply begin with the smaller representation characterized by S_0 and F_0? In general, however, the derivation of (S_0, F_0) from (S, F) is not trivial at all. The reason is that the invariant property e_0 does in general not correspond to a simple expression which is directly contained as a conjunctive factor in the expression of the state.

For example, in classical mechanics the primitive properties determining the state are the position and momentum variables. The conserved property of energy, however, depends on these variables in a rather complex way. The introduction of energy conservation will reduce the state space of the system to one of its constant energy hypersurfaces, but these surfaces have generally a far from trivial shape.

In general the introduction of a conservation principle will produce an *equivalence relation* on the state space, leading to equivalence classes of states characterized by a common value for the property which is to be conserved. State-transitions will only be allowed by the conservation principle if both states belong to the same equivalence class. In the technical language of mathematical physics these equivalence classes are called "irreducible representations" of the corresponding subgroup F_0 of operators for which the property is invariant.

The second type of dynamical constraints, the variation principles, will not introduce equivalence relations on state-transitions, but *order relations*. The basic idea is that even after all (practically computable) conservation principles were applied to reduce the set of operators and the corresponding set of states which can be reached from a given state, there still remains a rather large choice of possible transitions. These transitions will now be ordered according to some *evaluation criterion*. The variation principle states that that transition will actually take place which corresponds to the maximum of the ordering (or the minimum, depending on the convention used for computing the evaluation function).

Intuitively, this may be understood as a selection of the "best" transition among the available alternatives. This is basically an *optimization* problem. It requires that different states, trajectories or operators be mutually compared, according to a given evaluation criterion, so as to determine which one is the "best" for this criterion.

If the things to be evaluated and compared are the individual states, then the evaluation function is usually called a "potential" function. In this case every point s of the state space is accorded a numerical value P(s) which is called its "potential". The dynamical principle constraining possible evolution is that for a given time interval T the transition from an initial state s_i which will actually

occur is that which results in a final state s_t which has the lowest potential of all the states in a neighborhood $O(s_i)$:

$$\forall\, s \in O(s_i): P(s) > P(s_t).$$

If s_i is already a local minimum of the potential, then $s_i = s_t$, and no change of state will take place.

Remark that the use of a potential principle presupposes the existence of a continuity principle: if no continuity or neighborhood structure existed in the state space, then every initial state would be transformed immediately in that state of the space with the minimal potential (i.e. the global minimum of the potential function); all evolutions would stop immediately, and have the same result. This is clearly not a useful model of actual physical processes.

The local minima of the potential function are sometimes called "*attractors*": they indeed attract trajectories passing in their neighborhood towards their center. Once a system has reached an attractor state it remains there, it cannot evolve further.

For many systems a potential principle is not useful for representing their dynamical behavior. These are for example systems for which there are no attractors: their evolution never stops. The periodical movement of a planet around the sun is an example of such a never ending movement. In such a case it is meaningless to attribute absolute potential values to static states. Yet we can still use variation principles for comparing different trajectories.

The classic example of such a representation is the Hamilton principle in classical mechanics. Here it is supposed that two states, an initial state s_i (t_i) and a final state $s_f(t_f)$, are given, and that a trajectory connecting both:

$$\{s(t),\ t_i < t < t_f \text{ with } s(t_i) = s_i(t_i) \text{ and } s(t_f) = s_f(t_f)\ \}$$

is to be determined. All possible trajectories are evaluated by means of a function which is called the "action". The Hamilton principle now states that that trajectory will be actually selected by the system for which the action is minimal.

In general any system whose behavior can be expressed by a dynamical law or equation relating input variables, state variables and output variables as a function of time, can be equivalently represented as a decision-making or optimizing system (Mesarovic and Takahara, 1975). This means that all dynamical constraints governing the trajectory can be viewed as requirements that the trajectory be optimal with regard to a particular evaluation function.

In this study we have preferred the formulation with conservation or optimization principles to the more conventional formulation with dynamical laws or equations, because a dynamical equation does not give a very explicit analysis of how states, operators and time are coordinated in the determination of a trajectory.

Besides, the formulation with conservation or optimization principles is more general since it does not presuppose that you have complete knowledge of the dynamical constraints. In practice you often know only part of the conserved properties or ordering functions which allow you to select expected state transitions. In that case anticipation of the system's trajectory becomes a question of guessing or of heuristic search.

The classical representation frame, however, makes the supposition that the trajectory can be completely determined, without any uncertainty left. As we have already noted, this implies that the state be a complete characterization of the system at a certain time. Moreover it implies that the dynamical constraints (equation, conservation or variation principles) leave only one solution to the problem of which trajectory passes through a given initial state.

The billiard game example may help us to clarify some of the newly introduced concepts. Suppose we have a billiard game representation with predicates expressing besides the positions, also the colors of the balls. Clearly, during a regular game the colors of the balls will not change. Therefore, the property of the system expressed by the different color predicates is conserved, invariant or "a constant of the motion".

The operator group F characterizing this game might be the product of the translation group F_0 for the positions of the balls with the permutation group F_1 for the colors of the balls. Clearly we only need the subgroup F_0 for describing the possible state-transitions in the game.

In this example the reduction of the representation by the introduction of the principle of color conservation is trivial, because, variable (position) and invariant (color) properties can be easily factorized. This becomes less trivial if we use the dynamical representation with positions and momentums as primitive predicates. Here we might want to introduce the energy conservation principle. If there are no outside forces, the energy is purely kinetic and hence directly proportional to the sum of the squares of the momentums of the individual balls. The determination of a subspace of states and a subgroup of operators for which this property is invariant, is clearly not trivial.

For an example of a potential principle we may consider a kinematical representation of a billiard-game where the friction on the balls is so large that a moving ball on which no outside force is exerted stops almost immediately. This means that we can ignore all effects of inertia. In such a game the trajectories of the balls will be completely determined by the relief of the table on which they are moving. Gravitation will pull the balls towards lower regions while pushing them away from the higher regions. The lowest points of the table will function as attractors: in general, any ball will move towards one of its neighboring attractors, and it will stop its movement only when it has reached the attractor.

A simple example of a variation principle applied to trajectories instead of states can be seen in a regular billiard game (on a flat table-cloth). If you want a ball to move from one point of the table directly (i.e. without scattering) to another point, then what trajectory will be selected? Physical intuition tells us that the normal trajectory in the absence of outside forces is a straight line,

connecting initial and final positions. Yet a straight line can be characterized as the shortest path connecting two points. Therefore, the dynamical constraint determining the trajectory can be represented as an optimization principle with an evaluation function defined by the length of the proposed trajectory.

CHAPTER 5: The world view of the classical representation frame

5.1 Introduction

A basic principle of the cognitive paradigm is that perception is necessarily selective: not all information inherent in the environment can be processed in depth by the cognitive system. The adaptivity of the system is determined by its representation or cognitive structure, which selects those features of the environment which appear relevant to the general adaptation strategy of the system. These subjectively relevant features of the environment will be called the *world view* of the system.

It is clear that each representation structure will select a specific set of features of the world as focus of attention or awareness, while ignoring other features. The coordinated whole of these perceived features will form a kind of philosophy about how the world really is, how it evolves, and how it is related to the perceiving subject. We will now analyze the specific philosophy implied by the classical representation frame.

This will be subdivided in two sections: first, the *ontology* of the classical frame, i.e. the view of how things exist, independently of the observing subject; second, the *epistemology* of the classical frame, i.e. the view of how the subject can know these objective features of the world. The first philosophical domain corresponds to what we have called “physical science”, the second to “cognitive science”.

Traditionally, one distinguishes a third philosophical domain, *ethics*, which is the theory of how the subject should behave in the world. This, however, is the domain where the classical representation frame is the least explicit.

We can here only indicate the most general ethical presuppositions which can be derived from the classical representation. The *actions* of a subject can be represented with the same conceptual structures as the evolution of a physical system: each action corresponds to an operator which maps a given state of the world onto a new state. The sequence of allowed actions is constrained by certain principles, which can be variational or conservational.

Conservation principles correspond to rules of prohibition: certain features of the world must not be changed. For example, the moral principle “Thou shall not kill”, can be interpreted as a principle of conservation of the number of living human beings, constraining the sequence of actions of an individual.

Variation principles correspond to the establishment of moral values, of a distinction between good and evil, or, more subtly, of a distinction between better and worse. For example, the utilitarian ideal of the greatest happiness for the greatest number of people is a moral principle of the variational kind.

The whole of ethical constraints should in principle allow us to select one action as being the most appropriate, or the "right" one, for a given state of the world. There is however a basic difference between this ethical determinism and the ontological or physical determinism, implied by the classical frame. In both cases there is exactly one right solution to the problem of "which state-transition should occur next?"
In the physical domain, however, the transition which should occur is also the one which will occur. The choice we seemed to have between different operators to apply to the given initial state, is only apparent; the actual transition to take place was already determined beforehand.

In the ethical domain, the action which will take place is not necessarily the one which should take place. The acting subject is supposed to have a "free will", which allows him to actually choose between different alternatives, even though only one of these alternatives is considered optimal; he has always the possibility to make the wrong choice, i.e. to act "unethically" (on the other hand it is clearly impossible for a dynamical system to evolve "unphysically").

This problem of determinism is one of the areas where the classical world view appears to lack consistency.

5.2 The ontology of the classical frame

The most basic representation mechanism of the classical frame is the object-predicate scheme. It presupposes that the world can be cut up in a set of separate and invariant elements, to be represented as objects. The relations and interactions between those elements can then be represented by predicates, whose dynamics are represented by state-space trajectories constrained by conservation or variation principles.

The philosophy underlying the application of this scheme could be called "reductionism". It presupposes that the complex and changing world that we experience can somehow be reduced to a combination, according to predetermined rules, of a set of independent, unchangeable elements.

The problem with this view is that the elements of the representations (objects or predicates) usually do not correspond with clearly elementary outside phenomena. If we consider a "table" as an object of our representation, then it is obvious that we can replace this representation by a more detailed one, where the objects are a "table-top" and four "table-legs", which can be combined with the help of the two-place predicate "Supports":

Supports (table-leg 1, table-top)

would be a true proposition describing the more detailed structure of the table. Here we have two representations which both obey the classical framework, but which use different elements for describing the same phenomenon.

If we wish to have a consistent world view compatible with the classical frame, then we must suppose that these two representations are only approximations or abstractions of a more fundamental representation where the representation elements correspond to truly elementary, i.e. indivisible outside phenomena. These elementary phenomena are usually called "atoms" (in more recent formulations they are called "elementary particles").

This supposition implies that every phenomenon can be reduced in an unambiguous way to a combination of atoms. This means that every object in a representation which does not correspond to an atom, can be represented in a more elaborate way as a system consisting of several, more elementary objects, related through certain predicates.

Although in the coarse-grained representation the object is supposed to be invariant, in the fine-grained representation it can in principle change if the relations between its sub-objects are changed. For example, the object "table" can stop to exist if the table-legs are removed from the table-top, so that the proposition "Supports (table-leg 1, table-top)" is negated.

The requirement that the classical frame provide a complete and consistent representation of the world, implies that this is not possible for atoms: only the relations between atoms can change, the atoms themselves are permanent. Different atoms are distinguished by their properties (e.g. position in space) expressed by predicates. On the level of objects, where no predicates are introduced as yet, however, all atoms have the same function, hence are indistinguishable. This leads one to infer that different atoms are just distinct manifestations or instances of one and the same, unchanging essence or substance. This substance is called "matter".

If the completeness of the classical representation implies that there be elementary objects, it also implies that there be elementary predicates. An elementary predicate should give a complete characterization of an individual atom. However, the classical representations are not very explicit as to what are the fundamental properties of atoms. Depending on the theory, different sets of elementary properties are postulated (including e.g. mass, charge, position, momentum, energy,…), but one property which appears in all fundamental representations, is the position of the atom in the three-dimensional, Euclidean space.

This property is clearly fundamental in the sense that it is sufficient to distinguish any two arbitrarily chosen atoms: two different atoms can have the same energy, mass, charge or momentum, but they obviously cannot have the same position in space. This leads one to postulate "Absolute Euclidean Space" as a necessary component of the fundamental state space of an atom.

This absolute space concept has a role analogous to that of matter: the different elementary predicates of position can be conceived as representing "points", i.e. distinct instances or pieces of an invariant substance called "space". The difference between atoms, as elementary pieces of matter, and points, as elementary pieces of space, is that every atom is located in a particular space point, but not every space point contains an atom. This is because an atom

corresponds to an (invariant) object to which (variable) predicates, representing points, are attributed. Hence, an atom should have different potential points of location.

The classical scheme for representing change is a linearly ordered set of states. As we have seen, this ordered set can be indexed by a number system. In analogy with the space concept, the elements of this number system can be interpreted as denoting different elementary manifestations or pieces, of a universal substance which is called "absolute time". As in the case of space, the word "absolute" signifies here "invariant, the same for all atoms, objects or observing subjects". Hence, every event occurring to an atom is unambiguously characterized by a unique space point, and a unique "instant" or "time point". The absoluteness of the linear time ordering implies that it is always possible to establish objectively whether two events are simultaneous (i.e. occur at the same instant) or not.

Another feature of the world associated with this linearity is causality. The application of an operator f_T to a state $s_1(t_1)$ can be interpreted as a cause leading unambiguously to the effect: $s_2(t_1 + T) = f_T(s_1)$, i.e. the new state at a later time $t_1 + T$ resulting from the application of f_T to s_1.

In this conception of causality every cause has a unique effect:

$$\forall\, s_1 \in S, \exists\, !\, s_2 \in S: s_2 = f(s_1);$$

every effect has a unique cause:

$$\forall\, s_2 \in S, \exists\, !\, s_1 \in S: s_2 = f(s_2);$$

and the cause always precedes the effect. This corresponds to the idea that the whole of cause-effect relations constitutes a linear sequence, without loops or bifurcations.

In fact, in this framework the cause of a dynamic change is more than the sole initial state s_1; it also encompasses the whole of values of conserved properties or evaluation criteria that determine the dynamical constraints. These values (e.g. the potential energy of the state) usually represent external forces or influences which are not explicitly represented as objects, but which select the specific operator f to be applied to s_1.

The principle that for a given time interval T there is just one operator f_T transforming s_1 in s_2 implies that the trajectory of the state be completely determined. No ambiguity, uncertainty or freedom of choice is left. This feature of the classical world view is called *determinism*. It states that for a given initial state of the world all future states are already determined. There is no chance or randomness.

From this point of view change is only apparent; the fundamental structures which completely determine all events that occur, have occurred, or shall occur,

are absolute and permanent: matter, space, time, the dynamical laws are a priori fixed. Time should not be seen as a matrix of change or evolution, but as a mere additional dimension, to be coupled to the three dimensions of geometrical space for an unambiguous localization of all events.

This geometrical interpretation of time is reinforced by another feature of the classical world view: *reversibility*, or time-symmetry. The geometrical or topological structures of space are symmetrical: the relations between points "is in the neighborhood of", or "is at a distance d of", are symmetric. The linear order relation defining time, on the other hand, is antisymmetric by definition. In the classical frame, however, this antisymmetry is interpreted as a mere labeling convention or formality: a trajectory leading from $s_1(t)$ to $s_2(t + T)$ could as well have led from $s_2(t)$ to $s_1(t + T)$; the Hamilton principle determining the path from s_1 to s_2, produces the same result for the path from s_2 to s_1.

The idea that a state-trajectory or dynamical evolution could as well have been followed in the reversed order, can be traced back to two structural properties of the classical frame: the reversibility of operators, which is determined by the group structure, and the time-symmetry of dynamical constraints. The latter property requires that if a state sequence obeys a dynamical constraint, then also its inverse should obey the constraint. This requirement is implicit in all conservation principles, but not necessarily so for variation principles. In the case of a potential principle there is clearly no time-symmetry: a system can evolve from a state s_1 to an attractor state s_2, but it cannot move back to s_1, away from the attractor s_2. In classical mechanics, however, the variation principles are of the Hamilton type which is time-symmetric.

The world view implied by time-symmetric representations is one in which nothing can be really created or destroyed; every change can be undone. One way to express this property is to say that the information inherent in the state-of-the-world is conserved: the state itself may change, but its amount of information is invariant. In practice this means that the information inherent in a world-state s0 (t_0) is sufficient to compute all earlier states s (t_0-T) or later states s (t_0+T').

5.3 The epistemology of the classical frame

The function of the observing subject in the classical frame corresponds essentially to what we have called the "photographic camera view of the mind" (see sect. 1.3). This implies that the observer is fundamentally separate from the phenomena he observes; he does dot interfere with the outside world; he is merely passively receiving data which he tries to arrange in the correct manner so as to get a picture of the world which is as complete as possible.

True knowledge, in this view, is a kind of photographic picture or mirror image of the world. This corresponds to what we have called the isomorphic or denotative view of representation (sect. 2.5): the representation of the world that the knowing subject possesses is a structure isomorphic to the structure of the

world; its parts and elements (objects and predicates) denote separate parts and elements of the world (systems or atoms and their properties).

The correctness or "*truth*" of a representation in the classical frame is considered to be absolute: either the elements and relations of the representation are in a one-to-one correspondence with actual elements and relations of the world, or they are not; there is no ambiguity or subjectivity involved. This correctness can be checked in a simple way: it is supposed that for each predicate describing the state of a particular object there is an experiment or observation to be carried out on the phenomenon represented by the object which can have two results: "yes", the phenomenon actually has the property denoted by the predicate, or "no", the phenomenon has not got this property.

If the experiment is carried out well, in accordance with the intersubjective rules for the operational definition of the property, the answer is determined by the actual state of the object before the experiment began (cf. Aerts, 1983). Since the observer is supposed to be separate from the observed object, his observation should have no influence on the result of the test (however, it might have an effect on the state of the object after the test, see sect. 7.3). Hence, from the principle of determinism it follows that the result of the experiment, which establishes the truth or falsity of the predicative proposition, is absolutely determined, and could have been predicted with certainty by the observer if he knew the state of the system at the time the observation was made.

If the observer did not know the state, he could use this experimental procedure to determine it, by testing the truth or falsity of all predicates needed to define the state of the system. This would provide him with all the information needed for computing the future behavior of the system. This is another fundamental epistemological assumption of the classical frame: in principle any observer is able to gather complete information about a certain problem domain, so that all problems within this domain can be optimally solved by applying the deterministic inferential mechanisms of the classical frame. This feature of the knowing subject can be called "perfect or unlimited *rationality*" (Simon, 1957).

At the level of action, it implies that, given the ethical or dynamical constraints expressed as conservation or variation principles, the subject can always find the optimal sequence of actions in a given situation. This assumption forms the base of the classical theory of economics (Simon, 1957). The different presuppositions leading to the concept of classical rationality are made more explicit in the sciences of decision (Despontin, 1986; Roy, 1985).

6.1 Beyond the classical frame

In our search for a general characterization of adaptive representations, we have made a rather detailed investigation of a special type of representations, which were called "classical representations". Their analysis has revealed a quite elaborate structure, functioning as a general framework for the representation of many more specific problem domains. We have further analyzed how this structure determines the way a subject perceives the world and his interactions with it.

Many of the assumptions and implications of the classical frame appear intuitively evident. For example, the existence of objects, the reality of matter, space and time, the rules of logic and causality, the distinction between truth and falsity… all seem so natural that it appears difficult to imagine a world where these conceptions would not apply. Yet we must keep in mind that these concepts are only principles governing a particular representation of the world; they do not belong to the "objective reality" or the "Ding-an-Sich".

In this sense they are analogous to the Kantian categories, which are supposed to be necessary conditions of experience. The difference between the present viewpoint and the Kantian view however, is that these principles of experience and understanding are not supposed to be invariant and necessary a priori. We will now show that alternative principles exist for the organization of experience in adaptive representations.

These alternative representation principles will be found in several new scientific developments, where the complexity and variability of the phenomena studied makes it clear that you have to go beyond the classical frame if you want to understand them. As we have seen, the most general precondition for the efficacy of the classical frame is that the observed phenomena have a stable or invariant structure. It is to be expected then that phenomena characterized by instability, rapid change, or relativity will be difficult to fit into this framework. Since the normal evolution of science leads from simple, invariant phenomena towards more complex and changeful phenomena, it is clear that at some point in the development of a scientific discipline a crisis must occur resulting in a questioning of one or more of the classical presuppositions.

Since the classical framework as a whole appears so basic to our understanding of the world that it seems impossible to think without it, the number of presuppositions which are questioned and eventually replaced by alternative presuppositions, will tend to be minimal. Furthermore, it will be easier to change the presuppositions which are of a more technical nature (e.g. the group structure of the set of operators) than to change those which are really primitive (e.g. the existence of objects and predicates). Which particular assumption will be relinquished, will depend on the problem domain of the specific discipline.

For example, in the physical sciences one will tend to question ontological presuppositions, such as the absoluteness of space, whereas the cognitive sciences will witness discussions of epistemological assumptions, such as the absoluteness of truth.

Therefore, although there is presently a general critical attitude towards the classical frame, the alternatives proposed in different disciplines appear heterogeneous, incomplete, and often inconsistent. What is needed is an integrating framework where the different contributions are placed in the proper perspective with respect to the representation problem as a whole. We may hope that the adaptive representation concept, as it was developed until now, will provide the basis for such a framework.

However, in order to discuss the novel contributions in a more efficient way, we will first attempt to characterize classical representations in a more fundamental manner, so that the difference between classical and non-classical representation structures becomes more explicit.

6.2 Distinctions as basic elements of a representation

Possibly the primary function of a representation is classification. Any information-processing system divides the stimuli it receives into distinct classes, which correspond to distinct states of its internal representation. These states determine the further behavior of the system. The reason why a representation makes the search for an adequate reaction to a given stimulus more efficient, is because the classification of the stimuli into a number of classes which is smaller than the number of stimuli reduces the size of the search space.

If the number of states or classes would be as large as the number of physically different stimuli or situations the system may be confronted with, the mental search through the internal representation space would take as long as the actual search through the physical environment. This means that the system would be unable to make anticipations, since the internal computation of the effect of an external process would not be finished before the process itself is finished. In other words, if the representation of an environment would be as complex as the environment itself, it would become useless, since it would only duplicate this environment in a perfect way, without any gain in speed or simplicity.

Therefore, any process of perception implies a classification or distinction between phenomena. This classification operation has two aspects:

1) the phenomena which are put together in a class, are considered to be equivalent with respect to the system's goals, they are *assimilated*, they belong to the same equivalence class;

2) the phenomena corresponding to different classes are *distinguished* or discriminated, they belong to different equivalence classes.

The operations of *distinction*, and *assimilation* of phenomena necessarily go together. If a cognitive system would make no distinctions, only assimilations, it would be unable to perceive different phenomena, it would react to all situations in a uniform way; hence, it would be unable to adapt to a changing environment. On the other hand, as we have said, a system which would make no assimilations, only distinctions, would be unable to anticipate; hence it would also be unable to adapt. Henceforth, we will suppose that every distinction implies an assimilation, so that we can restrict our vocabulary to the one word “distinction”.

6.3 The invariance of distinctions in the classical frame

The fundamental mechanism of distinction comes out clearly in the analysis we made of the classical frame:

The use of objects as representational elements presupposes that we are able to distinguish between the object or system and its background or environment. At the same time it implies the assimilation of different states or appearances as belonging to the same object (cf. sect. 4.3).

A predicate clearly denotes a class of objects or phenomena (its extension) which are supposed to embody the same property (its intension) (cf. sect. 4.3).

The Boolean logic of propositions is merely an elaboration of this basic scheme of distinctions (cf. Spencer Brown, 1969). It is based on the operations of conjunction and disjunction which correspond to the intersection and union of the classes of phenomena denoted by the propositions, and on the operation of negation which corresponds to taking the complement of a class, i.e. the set of all phenomena which are not in the class, hence are supposed to be distinct from the phenomena within the class (cf. sect. 4.5).

The states are the propositions with the maximal degree of distinction; every other distinction between a proposition and its negation can be reduced to it, since, as we have seen, a state implies either the proposition or its negation. This means that the states correspond to the smallest equivalence classes, and that all other equivalence classes can be reconstructed as unions of state classes (cf. sect. 4.5).

Therefore, the logical part of a classical representation structure can be viewed as a static coordination of distinctions. The dynamical part of classical representations, on the other hand, is characterized by the requirement that all logical distinctions be conserved. Indeed, the basic mechanism for representing change is the replacement of a state expression by another state expression in which one (or more) of the primitive sub expressions has been negated. This means that the state has moved from one equivalence class into another one. The equivalence classes themselves and the distinction they imply, however, have remained invariant. Moreover, if two states are distinct and are subjected to the same operator, they remain distinct:

$$\forall s_1, s_2 \in S, \forall f \in F: s_1 \neq s_2 \Leftrightarrow f(s_1) \neq f(s_2)$$

(since all operations can be inverted the implication is valid in both directions). Equivalently:

$$s_1 = s_2 \Leftrightarrow f(s_1) = f(s_2)$$

(the equality sign "=" should not be read as absolute identity, but rather as equivalence of s_1 and s_2 with respect to the representation).

This corresponds to the classical conception of causality:

equal (equivalent) causes have equal (equivalent) effects;
distinct causes have distinct effects (cf. sect. 5.2).

It also corresponds to what we have called the *conservation of the amount of information*. Information was defined as that which allows us to make a selection (sect. 2.6). A selection, however, is nothing else than a distinction between those things which are accepted and those things which are rejected. The actualization of a representation state means that for every proposition it can be established whether the proposition is true or false (i.e. its negation is true). Therefore, for every distinction between a proposition and its negation one of both alternatives is selected. The more selections must be made for completely determining the state, the more information the state embodies.

In the classical frame, a change of state amounts to a different selection of alternatives. The number of alternatives, or distinctions to be considered, however, remains invariant. Therefore, the content of the information changes, but its amount remains the same.

The *conservation of distinctions* does not only apply to the effects of specific operators, but also to the choice of a particular operator determined by dynamical constraints. In the case of conservation principles it is clear that the selection criterion distinguishing between allowed operators and prohibited operators remains invariant. For variation principles, the basic distinction is that between "better" operators and "worse" operators. As we have seen, this distinction is supposed to be absolute in the classical frame (even though it depends on the state to which the operators are applied).

The structure of time is determined by two basic distinctions: that between *simultaneity* and *non-simultaneity*, and that between *past* and *future*. Both are supposed to be absolute: either an event x is simultaneous with an event y (i.e. it occurs at the same time), or it is not; if x is not simultaneous with y then either it happens before y or after y. This means that all observers, from all points of view, will make the same temporal distinctions.

Concluding, we may state that in the classical frame all distinctions, which determine the structure or form of the representation, are invariant: they remain the same for all times and for all points of view. The only thing which changes is the state, which determines the content of the information carried by the representation. We will now try to analyze how a number of alternative, non-classical representations transgress this rule of distinction invariance. In particular we will try to understand how they compensate for the loss of structural stability resulting from this loosening of distinction invariance.

CHAPTER 7: Quantum mechanics

7.1 The complementarity of representations

Quantum theory or quantum mechanics got his name from the assumed existence of a "*quantum of action*", i.e. a discrete unit of energy exchanged during microscopic interactions (i.e. processes at the scale of atoms or elementary particles). This assumption arose from the observation by Planck that electromagnetic radiation had to be exchanged in discrete units in order to explain the empirical laws derived for black body radiation. This idea was elaborated by Einstein who introduced the concept of a "photon" as a discrete, particle-like unit of electromagnetism.

Until then, in classical physics it was assumed that electromagnetism was carried by fields or waves, which spread continuously over space and time. This wave character could be demonstrated by interference experiments. Yet the same phenomenon which appeared to be continuous in these experiments, appeared to be continuous in other situations, exemplified by black body radiation and the photo-electric effect. Clearly, what everybody had seen as a wave, sometimes behaved as a particle.

The opposite phenomenon was discovered shortly thereafter. Electrons, which everybody thought to be particles, appeared to undergo interference, which is characteristic of waves. This double phenomenon, where particles behaved as waves, and waves behaved as particles, was called the "*wave-particle duality*".

In order to explain the signification of this strange result, we must situate the concepts of particle and wave in the classical frame, which completely determined the way of thinking of the physicists who were first confronted with these phenomena. As we have seen (sect. 5.2), a *particle* is an elementary instantiation of the concept of object. This means that it can be clearly distinguished from its surroundings, and that it has no apparent internal structure. Its definite position in (state) space evolves in a continuous manner, determined by dynamical constraints.

The concept of a *field* is an instantiation or materialization of certain of these dynamical constraints. It is defined by one or more numbers attached to all points in space. These field variables determine the evaluation function used in the application of variation principles. For example, in the application of the Hamilton principle in classical mechanics one uses the expression of the action, which contains the four components of the electric and magnetic potentials, which determine the electromagnetic field. If we want the trajectory of the particle to be continuous, it is clear that the evaluation function, and hence the field must be a continuous function of space.

A field is generally not constant in time; the values of the field at a given space point can vary, but only in a continuous manner. The (continuous) variations in time of a field, continuously extended over space, are called

"waves". A characteristic of waves is that they can be *superposed*: if two waves reach the same point, their collective effect is the sum of the effects of the two individual waves. E.g. if one wave makes a field variable at point x go up with an amount a, and the other wave makes it go up with an amount b, then the superposition of both waves will make it go up with an amount a + b.

In contrast to particles, waves cannot be distinguished from their surroundings. They do no occupy a determined position in space, they are spread out continuously over the whole of space. There is no boundary between wave and non-wave. On the other hand, particles cannot be superposed, because if two particles could be superposed, resulting in a third "superposition" particle, the original particles would lose their identity, they would no longer be distinguishable.

The problem with quantum effects is that we have to do with phenomena which appear to be both particles and waves! The statement that something is at the same time distinct from its surroundings, hence discrete or discontinuous, and continuous, is logically inconsistent. Therefore we are confronted with a paradox which demands a completely new way of looking at things.

The prevailing attitude towards this problem in physics, which is known as "the Copenhagen interpretation of quantum mechanics", is based on Bohr's *complementarity principle*. According to Bohr, we cannot know physical reality as it is, independently of ourselves. We can only make certain representations of the way we can interact with a physical phenomenon. These representations are necessarily formulated in the language of the classical frame:

> "it lies in the nature of physical observation, that all experience must ultimately be expressed in terms of classical concepts" (Bohr, as quoted by Jammer, 1974).

The necessity of using classical terminology when discussing observations followed for Bohr from our inability to relinquish our usual forms of perception (Jammer, 1974). In the case of the macroscopic entities studied by classical physics, such as planets or billiard balls, this creates no problems: their classical representation is well-defined, consistent, and deterministic. In the case of microscopic phenomena, such as electrons or photons, however, there is no complete and consistent classical representation: there are only partial representations (e.g. the wave representation and the particle representation) which are *complementary*.

The complementarity of representations signifies that, although they are mutually exclusive, they are jointly necessary—or complement each other—for an exhaustive description of the physical situation.

Bohr's reasoning in order to show the necessity of this complementarity of representations, is based on the role of the quantum of action in the physical *observation process* (Jammer, 1974; Bohr, 1958). Since there is an indivisible quantum of action, the energy and momentum exchanged during an elementary interaction is necessarily finite or discrete. Therefore, elementary processes are

characterized by a discontinuous change of state (e.g. from one discrete energy level of the interacting particle to another one). This phenomenon is sometimes called a "quantum jump".

The same principle must be applied to the observation process, which should be seen as an interaction between a (microscopic) object to be observed and a (macroscopic) measuring instrument, manipulated by the subject. Since the energy of the micro-object is of the same order of magnitude as the energy of the quantum of action exchanged during the observation process, the effect of this process on the state of the object cannot be neglected.

This would be no problem if it were possible to precisely determine what the effect is, that is to say in what way the observation has influenced the state. However, Bohr has shown by different thought experiments that no observation apparatus can completely determine this effect. The best one can do is to measure a certain feature of the process, e.g. the momentum transferred from the micro-object to the measuring instrument. The experimental arrangement needed for carrying out this measurement precludes, however, the determination of another feature of the object at the moment of observation, namely its position. Therefore, the experimental arrangements for determining position and momentum appear to be mutually exclusive: it is impossible to simultaneously determine both features of the object.

On the other hand we cannot gain any additional information by carrying out one experiment after the other one, since the first observation will perturb the state of the object in an undetermined way, so that the second observation will be unable to give information about the state before the first experiment was done. The measurements of position and momentum are said to be *incompatible*. Yet, according to Bohr, the representations of the phenomenon in terms of position, respectively momentum are complementary: both representations are needed for characterizing the relevant aspects of the phenomenon, but they cannot both be determined at the same time.

The complementary features of momentum and position correspond more or less to the wave nature, respectively to the particle nature of a micro-phenomenon. When such a phenomenon has a definite position, it can be said to behave like a particle. On the other hand, if it has a definite momentum, and hence an indefinite position, it behaves like a wave (with a wave-length inversely proportional to the momentum).

The impossibility of determining both position and momentum was expressed mathematically by Heisenberg, in his so-called "*indeterminacy relations*". These relations show in an explicit manner how a high precision in the determination of one of a couple of complementary variables implies a low precision in the determination of the other one: it is impossible to determine both with infinite precision.

7.2 The structure of the quantum formalism

We have started our discussion of quantum mechanics with the early interpretations of the quantum phenomena as they were experienced by the experimental

physicist. We shall now turn to the quantum mechanical formalism, that is to say to the structure of the explicitly constructed representation of these phenomena. However, we should add a remark here. To quote Jammer (1974):

> "This formalism […] was the outcome of a complicated process of trial and error and it is hardly an overstatement to say that it preceded its own interpretation, a development almost unique in the history of physical science".

Indeed, the Copenhagen interpretation, which was sketched in the preceding section, only developed some years after the first formalisms were proposed by Schrödinger ("wave mechanics") and by Heisenberg ("matrix mechanics"). These initial formalisms were synthesized by von Neumann, who formulated quantum mechanics as an operator calculus in Hilbert space. He thus laid the foundations for all subsequent quantum mechanical models.

The dynamical part of the quantum formalism is isomorphic to that of the classical frame: the evolution of a quantum system is represented as a time-parameterized sequence of state transitions, generated by an operator group, and uniquely determined by a dynamical constraint known as the "Schrödinger equation". The fundamental difference between classical and quantum models resides in the static, or logical part of the representation. This is expressed by a number of axioms, which determine the relation between the observations of the system, and its formal representation:

The states of the system are represented as vectors (or, more accurately, rays) in a *Hilbert space*. A Hilbert space is basically a vector space (which means that its elements can be added) with a positive inner product (which means that its elements can be multiplied, in such a way that the product of a vector with itself is a positive number). The addition of vectors corresponds to the "*superposition*" of quantum states. The quantum state characterized by the vector $s_1 + s_2$ is a superposition of the states characterized by the vectors s_1 and s_2 (in the sense of superposition of waves or wave functions. A wave function corresponds to a particular representation (Schrödinger or wave mechanics representation) of a quantum system).

This is a mathematical formulation of the general "*superposition principle*", which posits that for every two states of a quantum system you can find a third state which is a "superposition" of the other two states. This principle has no equivalent in the classical representation frame.

The physical quantities characterizing the system, and which can be measured by the observer, are called "*observables*", and are represented by self-adjoint linear operators acting on the Hilbert space.

According to the spectral theorem, every self-adjoint linear operator can be reduced to a linear combination of *projection operators*. These projections correspond to the observation of simple propositions or predicates about the system (i.e. variables with only two values: 1 and 0, or "yes" and "no"). In order

to simplify the discussion, we will restrict our attention to these binary observables. This does not diminish the generality of the analysis.

The observation process is expressed by a postulate which states that the possible measurement results for an observable A are the *eigenvalues* of the self-adjoint operator corresponding to A. For binary observables represented by projection operators the possible eigenvalues are 1 and 0: either the vector is projected onto its own ray, and then it is multiplied by 1, or it is projected onto an orthogonal subspace, which means that it is reduced to 0. This postulate is of a more technical nature.

More important on the conceptual level is the accompanying postulate stating that if a measurement yields a result corresponding to an eigenvalue a then the state of the system immediately after the measurement is an *eigenvector* corresponding to this eigenvalue. This postulate is sometimes called the *projection postulate*: during the observation process the state is projected onto the subspace, corresponding to the measured eigenvalue. This discontinuous state transition is also called "the *collapse* of the wave function": if a state is represented as a wave function defined in configuration space, the result of the measurement process will be viewed as a sudden, dramatic change of the shape of the wave, a "collapse".

For a system in the state s_1 the *probability* of finding an eigenvalue a as result of the measurement of the observable A is given by the projection of the (normalized) vector s_1 onto the subspace corresponding to a, multiplied with itself. This implies that if a state s_1 is already an eigenstate corresponding to a of A, then the probability of finding a is 1; the probability of finding any other result $b \neq a$ is 0.

In other words, the result of an observation is in general not determined, except when the state is an eigenstate of the observable. In conjunction with the projection postulate, this implies that if an observation of an observable A is repeated immediately afterwards, the result of the second observation will be the same as that of the first observation, since the state after the first observation was projected onto an eigenstate corresponding to the obtained measurement result.

For two different observations carried out one immediately after the other one, we must distinguish two different cases: either the operators representing the two observations *commute*, or they do not commute.

In the first case, the two operators can be shown to have a common set of eigenvectors. If the state of the system corresponds to one of those common eigenvectors, then both observables will have a determined result. Moreover the order in which the observations are carried out is irrelevant, since the projections of the state to the subspace of eigenvectors of both observations (projection postulate) commute.

In the second case, however, there are no common eigenvectors, so that there are no states for which both observables have a determined result. The

observables are said to be "incompatible". For example, the observables corresponding to the position respectively momentum variables do not commute, hence are incompatible. For such observables the order in which the observations are carried out will in general influence the results. The impossibility of getting determined results for both non-commuting observables can be expressed by indeterminacy relations of the Heisenberg type.

This Hilbert space formalism is in practice very complicated, so that it is difficult to see what are the fundamental characteristics which distinguish a quantum representation from a classical representation. It is nevertheless possible to simplify this formalism so that only the purely logical features of the representation are left for study. This is the domain of the approach known as "*quantum logic*" (Jammer, 1974).

It is founded on the result of Birkhoff and von Neumann (1935), who showed that a Hilbert space is completely determined by the ortho-complemented lattice of its subspaces. These subspaces correspond to projection operators which project onto a subspace. Therefore, any subspace corresponds to a binary variable or proposition about the system, whose truth or falsity can be established through an observation. (Such "yes-no" observations are called "questions" by Piron (1976)). In the classical frame the propositions form a Boolean lattice (sect. 4.5). The lattice formed by quantum propositions is not Boolean, however.

Usually the non-Boolean character of quantum representation lattices is expressed by stressing the fact that these lattices are non-distributive. Distributivity, however, is a rather technical feature which is based on the relation between the conjunction (meet) and disjunction (join) connectives in a lattice. In our reconstruction of the classical frame we have taken conjunction and negation as primitive connectives, while considering disjunction as a derived connective. Therefore we will look for another criterion for differentiating classical from quantum representation lattices.

The basic idea which must be expressed is that there is no longer a determined relation between a state s (i.e. an atom of the lattice) and a binary observable a (i.e. a general element of the lattice). If s is not an eigenstate of the observable a, then the result of measuring a is undetermined: different experiments will in general have different results, with a probability determined by the length of the projection of the vector corresponding to s on the subspace of eigenvectors corresponding to a.

This can be expressed by stating that the classical principle: either s implies a, or s implies the negation of a (i.e. the result of observing a is negative, the detected eigenvalue is 0), is no longer valid for quantum systems. Beside the eigenstates of a, which imply a or a', there is a third category of states for which the answer to the question "is a true or false?" is undetermined. These states correspond to the superpositions of states from the two other categories. Therefore, what the superposition principle really says, is that if you have two states s_1, s_2 which produce determined, but distinct results when a specific observation is performed upon them, then there always exists an intermediate

category of states for which the result is undetermined: in some cases these superposition states produce the same result as s_1, in the other cases they produce the same result as s_2.

This discussion can be further clarified by introducing a new relation between propositions or states: *orthogonality*. Two propositions or yes-no observables a and b are said to be orthogonal if any state which produces a positive result for a, will produce a negative result for b, and vice versa. In other words: a implies the negation of b, or, equivalently, b implies the negation of a.

" a is orthogonal to b" will be written: $a \perp b$.

The name of "orthogonality" comes from the fact that the subspaces corresponding to eigenvalue 1 of a, respectively b, are orthogonal (the product of vectors belonging to orthogonal subspaces is zero).

It can be shown that the set of states together with the orthogonality relation on this set completely determines the quantum lattice of propositions, and hence the Hilbert space (Finkelstein, 1979). The main argument of the proof goes as follows: for a subset A S of states, we can construct another subset:

$$A^{\perp} = \{s \in S: \forall\, s" \in A: s \perp s"\}$$

$A^{\perp}$ is the set of all states which are orthogonal to all states of A. This operation can be applied twice, so that we get $A^{\perp\perp}$, which is called the *orthogonal closure* of A. It has the following properties:

1) $A \subseteq A^{\perp\perp}$ (monotonicity)

2) $A^{\perp} = (A^{\perp\perp})^{\perp}$ (idempotence)

3) if $A \subseteq B$ then $A^{\perp\perp} \subseteq B^{\perp\perp}$ (inclusion preservation)

A set equal to its closure ($A = A^{\perp\perp}$) is called closed. Now it can be shown that the lattice of all closed subsets of the state space S is just the lattice of closed linear subspaces (corresponding to yes-no observables or propositions) of the Hilbert space (see Finkelstein, 1979).

Therefore, both lattice and Hilbert space are completely determined by the set S of all states together with its orthogonality relation $\perp$. This is a very simple structure compared to the original Hilbert space formalism with its vector space structure, positive inner product, operator algebra, etc. However, it is still more complicated than the basic logical structure of the classical frame.

Indeed, it is a well-known result for Boolean algebras that an atomic Boolean lattice is completely determined by its set of atoms (i.e. states). All other elements (i.e. propositions) of the lattice can be reduced to disjunctions or unions of atoms (i.e. states). Therefore, there is no orthogonality relation needed, only a set of states. Yet there is an orthogonality relation in the classical

frame, but at the level of states, it is trivial; all classical states are mutually orthogonal:

$$\forall\, s_1, s_2 \in S: s_1 \perp s_2 \textit{ iff } s_1 \neq s_2$$

(indeed, if s_1 is different from s_2, then s_1 implies the negation of s_2).

In the quantum case, however, not all states are mutually orthogonal: suppose s_1, $s_2 \in S$ such that $s_1 \perp s_2$, then you always can find a third state s_3, which is a superposition of s_1 and s_2, such that neither $s_3 \perp s_1$, nor $s_3 \perp s_2$.

This orthogonality relation also allows us to clarify the meaning of the superposition principle: superposition states are those states which are orthogonal to the same states as the states which were superposed:

$\forall\, s_1,s_2,s_3 \in S$: s_3 is a superposition of s_1 and s_2 *iff* $s_3 \perp s$, for all $s \in S$ such that $s \perp s_1$ and $s \perp s_2$.

In other words, the superposition states of a set $A = \{s_1, \ldots s_n\}$ of states, are those states which must be added to A in order to get its orthogonal closure $A^{\perp\perp}$ A.

This is the origin of the difference of classical and quantum representations: in classical representations, there are no superposition states, the orthogonal closure of any set is the set itself:

$$A^{\perp\perp} = A$$

Hence, all sets of classical states correspond to propositions or yes-no observables, whereas for quantum systems only a part of all state sets are orthogonally closed, hence correspond to yes-no observables.

Remark that the orthogonal complement $A^{\perp} = (A^{\perp\perp})^{\perp}$ of a state set A corresponds to the (physical) negation of the proposition determined by the orthogonal closure $A^{\perp\perp}$ of A. Indeed, if the state s of the system is such that the observation of $A^{\perp\perp}$ gives a determined answer "yes" (i.e. $s \in A^{\perp\perp}$), then the observation of A will by definition give a determined answer "no", and $A^{\perp}$ will be the largest such set.

7.3 A cognitive-systemic interpretation of quantum mechanics

We have analyzed the basic representation structure of quantum mechanics, and shown in which respect it differs from the classical structure. Moreover we have sketched the way quantum phenomena are traditionally interpreted. It is clear, however, that this interpretation is not very explicit, and that a lot of questions remain about the exact relation between the formalism and the phenomena it is supposed to represent. This incompleteness of understanding is illustrated by the

host of paradoxes and conceptual problems which can be found in books and papers about the foundations of quantum mechanics (see e.g. Jammer, 1974).

As we have remarked in sect. 6.1, the classical structures of logic, causality, space and time, are so deeply ingrained in our conception of the world, that it seems very difficult, if not impossible, to abandon them without falling into chaos and confusion. This is the reason why most of the quantum theorists, like Bohr, have stressed the necessity to maintain classical concepts for the description of physical observation, even though they acknowledged that these concepts did not apply to the microscopic phenomena to be observed.

This has led to a pragmatic attitude, which can be phrased as follows: "Why worry about the nature of quantum phenomena, since we are in principle unable to analyze them? The only thing which should interest us is the accuracy of the correspondence between the experimental results and the theoretical predictions. Since the accuracy of quantum mechanical predictions is very great, there is no real problem".

What is missing in this reasoning is the insight that a good representation should not only allow us to solve well-defined problems in an accurate way, it should also provide some heuristics or metaphors for the tackling of ill-defined problems, and thus provide a basis for an extension of its own domain of applicability. What is typical for the quantum mechanical representation is just the fact that, on the one hand, the basic formalism has not really changed since its appearance, more than half a century ago, while, on the other hand, none of the ill-defined conceptual problems connected to it, seem to have been solved in a satisfactory way.

For example, there still does not seem to be any satisfactory way to unify quantum mechanics and relativity theory, or to explain the experimental results connected to the Einstein-Podolsky-Rosen paradox (see sect. 8.6). Although no one would doubt the general accuracy of quantum mechanical predictions, such as those concerning the energy levels of the hydrogen atom, the fruitfulness of the quantum mechanical approach appears very questionable when applied to problems for which no explicit algorithm is proposed by the theory.

If we wish to find a more general representation of basic physical phenomena, then we must look further than the set of algorithms for the calculation of energy levels or probability amplitudes proposed by quantum mechanics. More specifically we must try to design a general philosophy or world view compatible with the quantum representation. This requires that we go further where Bohr stopped, namely at the insight that classical representations of our observation results are unable to give us a complete picture of the microphysical phenomena. The next step should be a search for a generalization of the classical frame.

In order to do this, a first requirement is that we be able to detach ourselves from the unconscious tendency to interpret everything through classical concepts. A first step towards this distancing or detachment, is the analysis of the classical frame we made in the previous chapters. It is just by making the general presuppositions of the classical frame more explicit, that it becomes

easier to conceive alternative frameworks where different presuppositions would apply. As we have seen, the most fundamental assumption of the classical frame, which can be considered to be the assumption which defines the "classicality" of representations, is the absolute invariance of distinctions.

Following this criterion, we see immediately why the quantum representation is not classical. The observation process, as formalized through the projection postulate, clearly does not conserve the distinction between quantum states, since different states can be projected onto the same eigenstate of an observable, whereas the same state can be projected onto two different eigenstates in two measurements of the same observable.

This accounts for the non-deterministic character of quantum observations: the knowledge of the initial state and of the operator representing the observable is insufficient to determine the result of the measuring process. The observation process is not causal in the classical sense, since equal causes (initial state + observation) can have different effects, whereas different causes can have equal effects. This is clearly in contradiction with the assumptions of classical physics (sect. 5.2 and 6.3).

The difference between quantum and classical observations is not one between observations which perturb the system, and observations which do not perturb the system, as is often thought. In classical physics it often happens that the measurement perturbs the system.

For example, to measure the mass of a billiard-ball, you could subject the ball to a known force during a known time interval, and measure the increase of velocity of the ball. This would allow you to compute the mass from the amount of transferred momentum. Clearly, this operation has perturbed the state of the ball.

Nevertheless, the process is causal, so that you can infer the characteristics of the initial state of the ball by observing the characteristics of the perturbed state, taking into account the known causal influence of your measuring apparatus upon the initial state. A causal process is basically a process which conserves distinctions, hence which conserves information. The original distinction between a large mass and a small mass for the ball has determined the distinction between a small increase of velocity and a large increase of velocity.

The only thing the observation operation has done, is to make this distinction more visible, to "*magnify*" it. We indeed cannot "see" that one ball is heavier than another one, but we can "see" that one ball moves faster than another one, after both were subjected to the same transfer of momentum. This is a general feature of measurement or observation processes in physics: to transform a "small" or "invisible" distinction into a "large" or "visible" distinction.

Let us consider an example from quantum physics: the difference between a spin up and a spin down state of an elementary particle is too "small" to be distinguished directly. Yet we can prepare a lot of particles in the same state and send them on a polarizer. If the state is spin up, the particles will pass the

polarizer, and a counter behind the polarizer will produce a signal perceivable by the human observer. If the state is spin-down, the particles will not pass the polarizer, and no signal will be produced. In this way the observer can unambiguously distinguish both states.

In this situation the information embodying the distinction between spin-up and spin-down has been transferred without perturbation from the system to the observer, even though the system itself has been perturbed drastically by the experiment. This is completely analogous to the classical case, where two different masses were distinguished.

However, there is a basic difference if more than one observation is considered. Although the distinction between spin-up and spin-down was transferred to the observer, other distinctions were neglected, i.e. assimilated by the measuring instrument: the polarizer clearly does not distinguish between particles with large and small momentums. In the classical case, the observer would determine a second property of the system by making another observation which would "magnify" the distinction between states which possess this property, and states which do not possess this property.

For example, in the billiard-ball situation, if the observer would like to determine the original momentum of the ball after he measured its mass, he can simply reason backwards from the detected velocity after the force was applied to the original momentum, taking into account the already detected mass. The reason he can do this is because the perturbation of the state was causal, hence conserved the initial distinction between large and small momentum.

In the quantum example, however, the initial distinction between particles with large and small momentums was lost when the particles were absorbed or scattered by the polarizer. The only thing we can infer by using the quantum formalism is that the particles which passed through the polarizer, were afterwards in an eigenstate corresponding to spin-up. But this does not give us any information about their momentum before or after the experiment. In so far as there was a distinction between particles with large or small momentums, this distinction has disappeared during the observation process.

In general we may conclude that during quantum observations certain distinctions are conserved, while other distinctions are lost. On the other hand, in classical observations all distinctions are in principle conserved, although not all distinctions are in general sufficiently "magnified" to attract the attention of the observer (for example, the observer will generally not notice a very slight difference in momentum between two balls).

A general condition for a quantum distinction to be conserved is that both states to be distinguished correspond to different eigenvalues of the observable which is measured. This implies that the states be orthogonal. States which are not orthogonal cannot be distinguished in a single experiment, they can only be distinguished statistically. Hence, the non-trivial orthogonality relation between quantum states is equivalent to the non-conservation of distinctions during quantum observations.

We must now look for the cause of this non-conservation. In order to make things more concrete we will analyze a *thought experiment*, used by Bohr for demonstrating the complementarity of position and momentum representations (Jammer, 1974).

Suppose we want to observe a particle by using a set-up consisting of a diaphragm with a slit. If the diaphragm is rigidly connected to a frame, the position of the particle can be determined (up to the, in principle, arbitrarily small width of the slit). However, the momentum exchanged between micro-object and diaphragm cannot be observed, because of the rigid connection of the diaphragm with the ideally unmovable (i.e. infinitely heavy) frame. Although there is an exchange of momentum, the change of state of the measuring instrument resulting from this momentum transfer is macroscopically indistinguishable. We could say that the difference between large and small momentum transfer is "assimilated" by the instrument; no distinction is left on the macroscopic scale of the "infinitely" heavy apparatus.

On the other hand we could distinguish between large and small momentum transfer by using a diaphragm connected to the frame by weak springs, so that it can move when hit by the particle. In that case, however, the information concerning the exact position of the passing micro-object is lost, because of the indeterminate location of the diaphragm with respect to the frame. Here the distinction between different positions of the particle is assimilated by the apparatus.

In these experiments one difference characterizing the interaction is magnified, hence made visible or distinguishable, whereas another one is "reduced" or "assimilated", hence made indistinguishable. Indeed, the change in velocity of the (heavy) frame resulting from the momentum transfer, is "infinitely" smaller than the change in velocity undergone by the (light) particle.

However, if we would reason classically, i.e. causally, we would remark that there still is a change of state of the apparatus caused by the interaction, and that it must somehow be possible to magnify this difference so that it becomes distinguishable. The problem here, however, is that to detect a difference, we must know exactly which state the apparatus was in at the moment the interaction took place. But this requires a determination of the microscopic state of a macroscopic apparatus.

The discrimination of properties that "small" demands a new experiment with a new macroscopic measuring instrument to magnify the microscopic differences characterizing the first instrument. Clearly the same limitations apply to this second experimental set-up, so that we need a third instrument for determining the microscopic differences registered by the second instrument. Obviously this leads to an infinite regression, where for each indeterminacy to be dissolved we need an extra instrument bringing itself new imprecisions into play.

Hence, there always remains a fundamental uncertainty as to the microscopic state of the macroscopic measuring apparatus, leading to the wiping out of certain distinctions characterizing the observed microscopic phenomenon.

The reason why this problem does not occur in classical physics is because the effect of microscopic differences between macroscopic instruments on the macroscopic interaction with another macroscopic object can be neglected. These "microscopic" differences are automatically assimilated, as well with respect to the state of the object as with respect to the state of the instrument. The only information needed for constructing a classical, causal representation of the macro-object, is that concerning macroscopic distinctions, i.e. distinctions at the same scale as those made by a macroscopic observation apparatus.

However, in the quantum realm, the distinctions needed for a causal determination of the micro-phenomenon are so fine-grained that they cannot be all magnified simultaneously by the observation instrument. This would not constitute a problem if different "magnifications" could be executed one after the other. Because of the quantum principle, however, the "indistinguishable" perturbation of the micro-state of the phenomenon by the interaction is of the same order of magnitude as the distinctions which are to be determined. Therefore, the distinctions which were assimilated inside the apparatus are also wiped out in the micro-object; therefore they are lost forever, and the result of subsequent distinctions to be performed on the object is basically indeterminate.

We may conclude this analysis by noticing that the fundamentally non-classical nature of quantum representation, which can be variously characterized by the concepts of indeterminacy, complementarity, superposition or non-trivial orthogonality, is simply an expression of the basic cognitive principle, stating that every distinction implies an assimilation. This principle is a direct consequence of another principle stating that a representation is always less complex than the thing it represents (cf. section 6.2).

One way to see the necessity of this principle is by noting that ultimately an adaptive representation should represent all possible interactions between the (inside) self and its (outside) environment. This implies that a representation to be really adaptive should also represent itself. Clearly this is only possible if representation entails simplification. Otherwise the part of the representation representing itself would be as complex as the whole, and this is obviously paradoxical.

Another way to express this idea is by introducing the principle of the impossibility of complete self-knowledge. To quote Finkelstein (1979):

> "[In this approach] Our inability to predict whether a photon from one polarizer will pass another is likened to the inability of a computer to predict whether a program will halt. Both seem like special cases of a general principle of self-ignorance, an anti-Solon principle (Solon said, "Know yourself!" The anti-Solon principle is that you cannot)."

In physics this principle is exemplified by the inability of a measuring instrument to determine its own microstate. In logic it is exemplified by the Gödel theorem, which entails that in general the completeness of a theory cannot be proven within the theory itself. We will now show how this principle can explain the non-classical features of the quantum formalism and its predictions.

7.4 From classical to quantum probability

Since the quantum representation is basically indeterministic, it can only make predictions which are of a statistical nature. These predictions express the probability of finding a certain result when an observable is performed on a given state.

They can basically be reduced to the transition probabilities P (s_f s_i), where s_i is the initial state in which the system was prepared, and s_f the final state onto which s_i is projected, and which is an eigenstate corresponding to the eigenvalue which was detected. P can also be interpreted as the *conditional probability* of finding the eigenvalue corresponding to s_f, knowing that the system is in state s_i.

It can be shown immediately that this probability is not classical. Classical probability theory, which was axiomatized by Kolmogorov, indeed presupposes the axiom of Bayes, which describes conditional probability:

$$P(a \mid b) = \frac{P(a \,.\, b)}{P(b)}$$

Here we see that the conditional probability P (a b) that a is true, knowing that b is the case, is proportional to the probability P (a.b) of a and b both being true. If we apply this formula to the quantum probability we find:

$$P(s_f \mid s_i) = \frac{P(s_i \,.\, s_f)}{P(s_i)} = 0$$

since s_i and s_f, being different states, can never simultaneously be true.

However, according to the quantum formalism P (s_f s_i) is only equal to 0 if s_i and s_f are orthogonal. In all other cases the application of the Bayes formula leads to a contradiction with the quantum formalism. In order to explain this result we must first analyze the basic assumptions connected with the probability concept.

The determinism of the classical frame implies that if an observer has complete information about the state of a system, then he can predict all results of observations of the system. However, in practical circumstances, the information the observer has got is often incomplete. *Statistical mechanics* is an extension of the classical frame, which aims to maintain as much as possible of the classical predictive mechanisms, without requiring that the initial information be complete.

The way this is achieved is by introducing the concept of a "*macrostate*" m, which represents the incomplete information received by an observer who is unable or unwilling to execute all observations or distinctions needed for the determination of the "*microstate*" s, which represents the complete information

determining the phenomenon. It is assumed that the microstate, which is "hidden" for the observer, obeys all rules and constraints of the classical frame. The macrostate, on the other hand, does not allow us to make determined predictions for all possible observations of the state. The principle holding for all microstates $s \in S$:

$\forall a \in E$: $s < a$, or $s < a'$,

where a is an arbitrary proposition about the system (cf. section 4.4), does not hold for macrostates $m \in M$.

The basic assumption of statistical mechanics is that every macrostate can be represented as an equivalence class of microstates, determined by an equivalence relation:

"cannot be distinguished (macroscopically) by an observer from".

The set M of macrostates then corresponds to a partition of the (micro)state space S. This implies that two distinct macrostates correspond to two *disjoint* sets of microstates.

The concept of *probability* can be fundamentally defined as:

"the number of positive cases divided by the total number of cases".

This applies as well to a series of *actual* events, in which some are considered positive, as to a set of *potential* events or states, in which some are positively distinguished. The first situation corresponds to the "frequency" definition of probability, where the probability of a particular type of events is established empirically by observing the frequency by which this type of event occurs in a controlled series.

The second situation is the theoretical or abstract representation of this phenomenon, which aims to provide mechanisms for predicting the empirical results from the first situation. As in all adaptive representations, this requires the specification of a space of potential events or states, from which certain events are to be selected. It is clear that the value of the probability defined in this way will depend on the number of potential events, i.e. on the size of the state space. This implies that representations with different state spaces will produce different probabilities.

However, in the classical frame it is assumed that for every system there is a unique set of (micro)states, which completely represent the information inherent in the system. The probability of an arbitrary proposition or macrostate will then be given by the cardinal number (more generally the measure) of the set of microstates which imply the proposition, divided by the cardinal number of the set S of all microstates:

$$P(a) = \frac{M(S_a)}{M(S)}$$

with: $S_a = s \in S: s < a$; M = measure upon S.

$$\text{Hence: } P(a.\,b) = \frac{M(S_a \cap S_b)}{M(S)}$$

since: $S_{a.b} = S_a \cap S_b$.

The formula for the conditional probability follows directly from these assumptions:

$$P(a \mid b) = \frac{M(S_a \cap S_b)}{M(S_b)} = \frac{P(a.\,b)}{P(b)}$$

(indeed, the probability of a, assuming that b is true, is equal to the number of positive cases $M(S_a\ S_b)$, i.e. the number of states for which a is true, and b is true, divided by the number of possible cases $M(S_b)$, i.e. the number of states for which b is already true).

Therefore the requirement of disjointness of macrostates implies that different macrostates be orthogonal:

$P(m_1 \mid m_2) = 0$ for $m_1, m_2 \in M$, $m_1 \neq m_2$.

Clearly then quantum states cannot be modelled by classical macrostates. Yet, many people have tried to explain the indeterminacy of the quantum representation by assuming the existence of "*hidden variables*", i.e. classical microstates, which would determine the observation results in a causal, deterministic way, but which would not be known by the observer. The impossibility of reconstructing a quantum representation starting from classical representations with hidden variables was proven by von Neumann, and from a different viewpoint, by Bell (cf. Jammer, 1974, and sect. 8.6). Both demonstrations, however, require certain assumptions, which amount to the principle that the classical macrostates be disjoint.

If this principle is relinquished, however, it becomes possible to introduce macrostates, $m \in M$, corresponding to generally *non-disjoint sets of microstates*, which behave like quantum states, i.e. which have a non-trivial orthogonality relation defined by:

$m_1 \perp m_2$ *iff* $S_{m1}\ S_{m2} = \varnothing$

The microstates s_i which are elements of the sets S_m are called "infrastates" by Finkelstein (1979). The probability of transition between macrostates could then be defined by:

$$P(m_1 \mid m_2) = \frac{M(S_{m1} \cap S_{m2})}{M(S_{m2})}$$

7.5 Information transfer during the quantum observation process

In order to explain the physical meaning of this formal construction, it is necessary to give a concrete interpretation to these microstates which have the strange property that the same microstate can imply different macrostates. One possible point of view is to see a microstate as a complete characterization of the microscopic relation between the observed object and the part of the measuring instrument with which it interacts directly. This means that if we would know the microstate, then we would be able to predict unambiguously the outcome of the experiment, i.e. the value measured and the state of the particle after the measurement.

However, as we have discussed, only a fraction of this information inherent in the microstate is sufficiently magnified by the remaining part of the measuring instrument to become macroscopically distinguishable. This fraction corresponds to the macrostate. Hence, certain small differences between microstates will be assimilated, so that every macrostate includes different microstates. However, which fraction of the information is made distinguishable will depend on the macroscopic arrangement of the measuring instrument, which determines the way the microscopic event of interaction is *coupled* to the macroscopically visible signal to the observer.

For example, in Bohr's gedankenexperiment with the diaphragm, there are two different manners of coupling the diaphragm (i.e. the directly interacting part of the instrument) with the frame (i.e. the part of the instrument which "magnifies" the change of state undergone by the other part): either the coupling is rigid, and then the information about the position of the interaction is accurately magnified, or the coupling is flexible, which allows us to magnify the information about the momentum exchanged during the interaction.

Hence, different ways of coupling will lead to different macrostates, even though the microstates of the interaction can be the same. On the other hand, the macrostate characterized by determined momentum, and the one characterized by determined position correspond to incompatible, or complementary, arrangements, and hence it is impossible to give an operational meaning to the conjunction of both states, even though formally their corresponding sets of microstates have a non-empty intersection.

This complementarity of macroscopic couplings can also be viewed as a complementarity of *subject-object distinctions*. Indeed a representation is essentially a scheme for making distinctions, and the most basic of these

distinctions is the one between subject and object (see section 2.1). As we have shown by two examples, however, the distinction is not absolute or invariant.

The same kind of distinction is made in physics between the object and the observing subject. Classically this distinction is supposed to be absolute: observer and observed are essentially separate (see section 5.3). In quantum mechanics, however, this distinction is blurred.

This can be illustrated by the *Schrödinger's cat paradox*, where the life or death of a cat depends on the result of a yes-no observation of a quantum phenomenon. This phenomenon is supposed to be in a superposition state such that the result "yes" (which does not interfere with the life of the cat) and the result "no" (which triggers a mechanism which kills the cat) have equal probability. According to the quantum formalism, the observation of the phenomenon leads to a collapse of the wave function, which results, in one case out of two, in the killing of the cat.

The paradoxical question is then: who has killed the cat? In other words, which event in the sequence of events which form the observation process, has caused the collapse, and hence the death of the cat: the microscopic interaction between the quantum phenomenon and the measuring instrument, the magnification of this event by the rest of the instrument, the perception of this magnified signal by the observer, or the awareness of the observer that this signal signifies the collapse of the wave function?

That this problem is far from trivial can easily be seen by imagining the following situation: the experimental set-up is prepared by a physicist A at time t_1, the interaction leading to a macroscopic signal happens at a later time t_2, the registered signal is noticed by a second physicist B at time t_3, who is told at time t_4 by A that this signal means that the wave function has collapsed and hence that the cat is dead. At what time did the cat die: t_1, t_2, t_3 or t_4? And who killed it: A, B or the measuring instrument?

From our point of view the wave function collapses when the information from the microstate is magnified, and thus creates a "collapsed" macrostate. The process of magnification, however, will depend on what we have called the "coupling" between microscopic interaction and macroscopic observation frame.

The coupling really determines the *boundary*, distinction, or interface between the autonomous phenomenon, and the macroscopic observation instrument controlled by the subject. This boundary can be conceived as a filter, or a semi-permeable membrane, which only allows a certain type of information to pass (cf. sections 9.3 to 9.5). The "magnified" information, forming the macrostate, should then be seen as the information which passed the filter.

The coupling, and hence the filtering, will in general be different for different types of observations. The specific type of coupling used during an observation is however not represented in the quantum formalism: the projection postulate does not specify how, where, or when the collapse of the wave function occurs. This ambiguity leads to the paradoxes of the measure-

ment process (e.g. Schrödinger's cat, Wigner's friend, the Zeno paradox, …). The solution of these paradoxes, and more generally a better understanding of the quantum representation, requires the construction of more explicit models of the .

A nice example of such a model is given by Aerts (1986). He considers an arrangement for measuring the spin of a particle, which consists of a sphere with two charged poles, which exert an attractive force on a third charge representing the spin. If the spin is attracted by the upper pole, the measurement result is "spin up". If it is attracted by the lower pole, the result is "spin down".

This result will depend on two parameters: the difference of charge q between upper and lower pole, and the distance or the angle A between the spin and the upper pole. The larger q, and the smaller A, the larger the force exerted by the upper pole upon the spin. Given those two parameters, the model is deterministic: the result of the measurement is determined by the relative strength of the two forces exerted upon the spin, and this strength can be computed classically.

However, Aerts assumes that we lack information about the relative charge q. By supposing that q can vary within a determined interval, and that all values within that interval have equal probability, he derives the algebraic expression for the probability of getting the result "spin up" for an initial state characterized by a given angle A. This probability expression coincides perfectly with the expression derived from the quantum representation of spin. The purpose of the argument is to show that a quantum, i.e. non-classical expression for probability can be derived by supposing that there is a classical lack of knowledge on the microstate of the instrument.

However, I wish to make two remarks about this argument. First, we do not need the exact microstates of the spin and the instrument (represented by the two parameters q and A) to determine the observation result. It is sufficient to know the relation between both parameters, which determines the strength of the attractive force on the spin in the upward or downward direction. This relation can be represented by a "relational microstate", which may be called the "infrastate". This infrastate is less specific, or contains less information than the conjunction of both microstates represented by q and A.

Different observation results, leading to different macrostates can now result from equal infrastates. Indeed, different macroscopic arrangements for measuring spin in different directions, can still be in the same microscopic relation with the object they are measuring. This allows us to recover the non-trivial orthogonality relation and transition probabilities between macrostates, where the probability of a transition from m_2 to m_1 is proportional to the number of infrastates common to m_1 and m_2.

The second remark I want to make about the Aerts model is that it is not as classical as it seems. Although the explicit derivation of the attractive force is purely classical, it is implicitly assumed that the position of the attracting pole corresponds to a fixed "attractor state", to which the spin will stick after it was attracted. The existence of attractor states is, however, not compatible with the

classical principle of reversibility (see section 5.2). This irreversible attraction implies that after the measurement the information about the initial angle A is lost: all subsequent observation results will only depend upon the position to which the spin was attracted, not upon its initial position before the attraction.

This explains why we do not want to include the information about A in the infrastate. Hence, it is not necessary to attribute both a (determinate) value to A and an (indeterminate) value to q. The only thing which counts is the relation between A and q, which is microscopically determinate, but macroscopically probabilistic. The non-classical (i.e. non-Bayesian) structure of this probability is due not so much to the fact that the state of the micro-object (A) is determinate, whereas the state of the instrument (q) is not, as Aerts proposes, but to the fact that the probabilistic macrostates corresponds to overlapping sets of infrastates.

8.1 The relativity of reference frames

As we remarked earlier (sections 1.1 and 3.3), to construct a useful representation of a system we need more than an abstract set of states: we need an indexation scheme for labeling the states in such a way that their operational meaning can be unambiguously specified. In mechanics such a scheme is often called a "*reference frame*". It allows us to express the time and position of an object in relation or in reference to a system of axes representing the viewpoint of the observer.

The numbers expressing the distance between the origin of the system, which represents the position of the observer, and the point on the axis which is the projection in parallel with the other axes of the position of the object, are called the spatial *coordinates* of the object. These coordinates unambiguously determine the position of the object. In order to determine them operationally we need a *measuring unit* for measuring the distances (i.e. a rigid *rod)*, and a frame for specifying the directions of the axes. Time, on the other hand, is a coordinate measured by a *clock* carried by the observer, which labels subsequent states or events, independently of their localization with respect to the spatial reference frame.

In the classical frame once the measuring units (rods and clocks) are fixed, the measurement of length (spatial) and duration (temporal) is supposed to be invariant. By changing the reference frame we may change the coordinates of an event, determining its position in space, its instant in time and its velocity (i.e. the relation between spatial and temporal coordinates), but we will not change the temporal or spatial relations between events, i.e. their relative distance or duration. In particular, if two events are simultaneous (i.e. their relative duration is zero), then they remain simultaneous for all observers and reference frames.

This is no longer true in *relativity theory*. This theory is founded on two postulates: the *principle of relativity*, and the *invariance of the speed of light*.

The first principle states that physical laws (i.e. dynamical constraints) should have the same form for observers using different (inertial) reference frames. You could see this requirement as a direct consequence of classical determinism: there is only one solution to the problem of how a physical system evolves; hence different observers should find the same solution. You could also see it as a requirement that the representation of the way states change (dynamical constraints) should be more invariant than the representation of their position in space and time. Hence, the principle implies that the specific indexation schemes determining spatial and temporal coordinates (and hence velocities) are relative; only the dynamical laws can be absolute. In order to express this principle mathematically you need a group of transformations,

which transform the states belonging to one frame into the states belonging to the other frame, but in such a way that the physical laws remain the same.

The fundamental laws of mechanics are those formulated by Newton. They imply that a massive object on which no force is exerted, moves at a constant velocity. The requirement that this law be literally valid in the frame we are using, singles out a special class of frames; these are called inertial frames. For an observer using such a frame the movement of a force-free particle is not accelerated. The relativity principle then states that the three laws of Newton should be valid in all inertial frames.

Remark that this principle does not give us any information about physical phenomena. Its only function is to coordinate different representations. Indeed, if we would interpret the principle physically, we would conclude that Newton's laws are valid in all those frames which were chosen such that Newton's laws would be valid. This is obviously tautological.

The interesting thing about the principle is that it allows you to determine the form of an actual transformation group relating different frames. By assuming that the reference frames are Cartesian and isotropic, it can be shown that it leads to two distinct types of transformations (Rindler, 1977).

The first type forms what is called the *Galilei group*. This is the group used in classical mechanics. The second solution to the problem is a transformation group determined by an indeterminate parameter V^2.

In order to explain the physical meaning of this parameter, we must make an extra assumption, based on the following dichotomy: either particles can be accelerated to arbitrarily large speeds, or they cannot. Because of the continuity principle, which excludes movements with infinite speed (see section 4.6), we shall suppose that they cannot. Then there must exist a least upper bound $\in$ to particle speeds in any one inertial frame. Because of the relativity principle, this maximal speed c, which determines a physical law, must be the same for all inertial frames. This implies that a transformation from one inertial frame to another one must conserve this speed. If we apply this requirement to the transformation groups we already found, then there remains only one solution, namely the one with $V = c$.

This implies that the Galilei transformations are ruled out by the assumption that there exist an invariant maximum speed. The remaining transformations characterized by the parameter c are called the *Lorentz transformations*.

The existence of an invariant speed c is called the second postulate of relativity theory. In this case the principle seems to have a physical meaning, and it is possible to do experiments to determine the magnitude of c. Empirically c appears to be the speed of light (or of electromagnetic radiation in general). The second postulate was confirmed by the Michelson-Morley experiment, and by the insight that Maxwell's laws of electromagnetism are not invariant under the Galilei-transformations, but under transformations of the Lorentz type.

The switch from Galilei transformations to Lorentz transformations had profound consequences for the structure and the world view of the mechanical representation frame. One of the more spectacular consequences is that in the novel representation the length of rods and the frequency of clocks should

diminish whereas their mass should increase, when the rods and clocks are accelerated to a speed near to c.

More important on the conceptual level, however, is the new relation between space and time: in the relativistic frame they are no longer separate, but form one indivisible whole, "space-time".

8.2 The relativity of simultaneity and synchronization

In the Galilei transformations the time coordinate of an inertial frame does not transform, only the space coordinates change. Hence, time is invariant or absolute. If two events happen at the same instant of time in one frame, they are simultaneous in all frames.

In relativity theory, however, the Lorentz transformations mix the time and the space coordinates in a way dependent upon the relative velocity between the initial and the transformed frame. Hence, frames with different velocities will have different time coordinates. If two events are simultaneous in time, but separate in space in one frame, they will in general no longer be simultaneous after a transformation. Since the new time coordinate depends on the previous space coordinates, the events with different space coordinates will now also get different time coordinates.

What we see here is that the distinction between simultaneity and non-simultaneity becomes variable or relative. It may even get worse, and we may find events a and b so that a happens later than b in one frame (a's time coordinate is larger than b's coordinate), whereas b happens later than a in another frame. Hence, also the distinction between past and future loses its invariance! However, as we shall see later (sect. 8.3), there are constraints which prohibit the interchange of past and future in many cases.

Let us try to understand on a more physical level why it is impossible to establish absolute simultaneity. If I want to send a message to another observer, who is separated from me in space, then I must use a physical *carrier* for this information. But all physical carriers (particles, macroscopic objects, electro-magnetic waves) are subject to the existence of a limit speed c. Hence, *I cannot transmit information instantaneously*; a finite time must necessarily elapse before the signal can reach its destination.

This principle is sometimes called the principle of *locality* or *causality*:

I cannot cause an event which is separated from me in space but not in time; the only phenomena I can influence instantaneously are those with whom I have local contact, i.e. which are not spatially separated from me.

The way this principle was formulated was with the emphasis on the sending of information. However, it also applies to the receiving of information. Perception or observation is basically a process whereby a subject receives information from an object. Suppose that the object undergoes a certain state transition. Call

this event e_1. Suppose that e_1 causes a perturbation in a physical carrier (e.g. the electromagnetic field) and that this perturbation propagates to the observing subject. There it induces a state-transition in the subjects perceptual system (e.g. an excitation of the optical nerve). Call this perceptual event e_2. Clearly e_1 and e_2 cannot be simultaneous, if object and subject are spatially separate; e_1 necessarily precedes e_2:

$e_1 \; P \; e_2$.

Suppose the observer reacts immediately and sends a signal back to the object, which causes a state-transition e_1". Clearly:

$e_2 \; P \; e_1$", and hence by transitivity: $e_1 \; P \; e_1$".

Suppose that both signals travel with the maximum velocity c. Even then there is a non-zero time interval T between e_1 and e_1". During this time interval many events or state-transitions will have occurred to the object. Now, since e_2 is temporally in between e_1 and e_1", one of these intermediate events should be simultaneous with e_2, according to the classical conception of time. However, it is operationally impossible to determine which one, since it is impossible to connect e_2 in a causal way to one of these intermediate events (cf. Reichenbach, 1958).

We can only define simultaneity in a conventional way, by calling two events simultaneous if two clocks which are locally connected to the events indicate the same time when the events occur. However, the time indicated by a clock depends on the way the clock is *set*, i.e. on the moment which is taken as the zero of the measurement scale. If two different clocks are to give consistent results, they should be *synchronized*, i.e. the zeros of their time scales should be made to coincide.

This is no problem if both clocks have the same position in space. However, if they are spatially separated, the synchronization requires that a signal be exchanged between the clocks. The transmission of this signal requires a finite time, depending on its velocity. But velocity is defined as distance divided by duration, and the determination of the duration of a transmission requires that it is possible to determine the time of departure *and* the time of arrival of the signal. In order to determine these two times in an absolute way you need two clocks which are already synchronized.

Hence, we come to a circular reasoning: you can only synchronize two clocks in an absolute way, if they are already synchronized!

There are two ways to evade this circularity: either you assume that it is possible to send signals with unlimited speed, and then you can neglect the finite duration, or you make an arbitrary convention (e.g. that the speed of the signal has a certain fixed value c) which allows you to determine the duration of

the transmission. The first approach is implicit in the classical frame. The second approach is used in Einstein's theory of special relativity.

It should be remarked that the standard synchronization used in special relativity, which is such that the velocity of light is the same in all directions of movement is not physically necessary. Einstein's definition is maybe the one which leads to the simplest model, but it is not the only possible one (cf. Reichenbach, 1958; Sjödin, 1979). A property of the Einsteinian synchronization is that reference frames which are in relative motion, will attribute simultaneity to different pairs of events. This, however, should not be taken as a proof of the relativity of simultaneity, as is often done. It is always possible to define a (non-standard) synchronization such that the simultaneity in a reference frame K would be the same as that in another frame K' which is in motion relative to K. The relativity of simultaneity is not a consequence of the relativity of (inertial) motion, but of the existence of a finite limiting speed for causal propagation (cf. Reichenbach, 1958, p. 146).

A second remark we could make, is that the impossibility to establish the absolute time at which a distant event occurred, is similar to the impossibility to establish the microstate of a quantum system (cf. section 7.5). In both cases we have an observer and an object, which are spatially separated, but which exchange information. Both quantum mechanics and relativity theory can be interpreted as theories which represent restrictions upon the exchange of information. Hence, what both theories have added to the framework of classical physics, has not to do with the material or energetic properties of physical processes, but with their informational properties. It is the inclusion of these informational principles which leads to the direct connection between physical and cognitive problem domains.

In the relativistic case the constraint on information transfer is that its velocity is limited. This entails a limitation on the possibility of distinguishing between simultaneous and non-simultaneous events. In the quantum case, the limitation is on the type of information which can be transferred through a specific coupling between object and observation apparatus. This also entails a limitation on the possibility of making distinctions, which is represented by the non-trivial orthogonality relation. As we will show later (section 8.4), both restrictions can be seen as consequences of a general principle of the impossibility of circular information transfer, leading to the impossibility of perfect self-determination.

In order to see the analogy with the quantum problem of impossible self-knowledge (section 7.5.) we could imagine the following situation: if I want to observe myself, I need a mirror, which is spatially separated from my body. However, the light which is reflected in the mirror needs a finite time to travel from my body to the reflecting surface and back to my eyes. During this time interval my physical appearance will have changed, however slightly. Hence, the image I see can never be a completely faithful representation of the actual state of my body.

We have seen that the existence of an invariant limiting speed c entails the relativity of simultaneity. However, this relativity implies the relativity of synchronization and hence of the determination of velocities. Then how can we speak about an invariant velocity, if there is no absolute way to measure velocities? If we say that the velocity of light is invariant then we really mean that the velocity is independent of the velocity of the sender or the receiver of the light signal.

This is typical for "*wave-like*" signals, in contrast to "particle-like" or "ballistic" signals. If I perturb a certain medium, e.g. water or air, then this perturbation propagates away from me with a velocity determined by the properties of the medium (e.g. its density). There is no way I can accelerate or decelerate this propagation by "pulling" or "pushing" on the perturbation.

On the other hand If I shoot a bullet out of a rifle, the velocity of this bullet with respect to the air will be greater if I am moving in the same direction, i.e. if I am "pushing" the rifle, and smaller if I am moving backwards. This type of propagation is called *ballistic*.

There is, however, one difference between light waves and waves in material media (e.g. sound): the relative velocity of the sound wave will depend on the relative velocity of the medium with respect to the observer. This does not seem to be the case for light signals. The Michelson-Morley experiment has shown that the velocity of light is independent of the relative velocity of the observer with respect to its supposed medium, the "ether". This led Einstein to reject the existence of such an absolute and pervasive medium.

We still must answer the question how the velocity determination can be made independent of the synchronization. The synchronization problem arises because we use two spatially separated clocks. This can be avoided by using a mirror which reflects the light back to its source, so that both the time of arrival t_2 and the time of departure t_1 can be measured by the same clock. The velocity v could then be put equal to the double of the distance l between source and mirror, divided by the time interval between departure and arrival:

$$v = 2l / (t_2 - t_1)$$

Remark, however, that we assume here that the velocity of the signal is the same for both directions: from source to mirror, and from mirror to source. This corresponds to the standard synchronization used in relativity theory: according to this assumption, when the signal arrives at the mirror a clock situated there should be set to a time:

$$t_m = (t_2 + t_1) / 2$$

With this synchronization the two *one-way velocities* v_l (from the source to the mirror) and v_r (from the mirror back to the source) are equal to the average *two-way velocity* v:

$$v_l = l / (t_m - t_1) = v_r = l / (t_2 - t_m) = v$$

For different synchronizations (where a different value is given to t_m) v_l and v_r will not be equal.

In general, the only thing we can absolutely determine is that the "average" velocity v for both directions has an invariant value v = c, which is independent of the orientation of the set-up and of the relative velocity of the mirror with respect to the source.

Suppose now that it would be possible to send signals with an average, two-way velocity larger than c. Such faster-than-light signals are usually called *tachyons*. We could then synchronize two clocks, one at the source of the signal, and one at a reflecting mirror, in such a way that the velocity of the tachyon would be the same in both directions. This synchronization will in general be different from the standard synchronization for which the velocity of light is the same in both directions.

It can be shown that for one particular reference frame, both synchronizations will coincide. The proof can be found in (Sjödin and Heylighen, 1985), for the case where the tachyons have no maximum limiting speed.

The existence of such a preferred frame would, however, be in contradiction with the principle of relativity, which entails that it is impossible to distinguish two inertial frames by doing experiments in them. Since both the invariance of the speed of light, and the relativity principle seem to be experimentally well-validated, this argument can be used to exclude the existence of tachyonic signals.

8.3 The invariance of the causal structure of space-time

This leaves us with two kinds of signals:
1) signals with an invariant velocity of c;
2) signals with a relative or variable velocity which is less than c.

The existence of these two types of causal propagation can be used to structure the set M of all possible events. In relativity theory an *event* is supposed to be a change which is instantaneous, and of atomic dimensions. A typical example is the scattering of two elementary particles. Such events can be represented by a point in space, denoting its position, and a point in time, denoting the instant it occurred.

However, because of the relativity of simultaneity, it becomes impossible to attribute the same point in time to two events, at a different point in space. Equivalently, there is no absolute way to attribute the same point in space to events happening at different points in time, because different observers would in general attribute different positions to the second event, even if they had attributed the same position to the first event. This makes it meaningless to attribute independent spatial and temporal positions to events; the only thing

which can be determined in an invariant way is a combined "*spatio-temporal*" set of events.

In order to structure such a set we need a number of invariant relations between events. These are provided by the different possibilities of exchanging information, such that the departure of a signal corresponds to an event a, and its arrival to an event b:

If the signal is of the light-like type, the relation between a and b is called "*horismotic precedence*" (HP).

If the signal is of the ballistic-type, the relation is called "*chronological precedence*" (CP).

The union of both relations: "it is possible to send a signal (light-like or ballistic) from a to b" is called "*causal precedence*", or "precedence" for short (P).

The basic structures of space-time, as represented in relativity theory (special and general), can be derived from the set of events provided with these three relations (Zeeman, 1964; Kronheimer and Penrose, 1967; Woodhouse, 1973).

The precedence relation can be interpreted as the fundamental temporal order:

a P b signifies that "a is able to cause b ", or that "b happens after a".

Remark that this relation is a *partial order*; it is not complete, or linear, like the classical temporal order (section 4.6). This means that for two arbitrary events a and b, we cannot establish in an absolute way that either a P b, or b P a (or a simultaneous to b).

In this sense, the innovation of the representation structure brought by relativity theory can be reduced to the one structural feature, that the classical relation of precedence loses its completeness. This is analogous to the way we analyzed the quantum mechanical revolution in representation. Here the innovation can be reduced to the fact that the orthogonality relation between states loses its completeness. Whereas the relations of precedence and orthogonality had a trivial structure in the classical frame, in relativity theory, respectively quantum theory, these relations play an essential role in the determination of the representation structure. In quantum mechanics the orthogonality relation determines the Hilbert space structure. In relativity theory the precedence relation(s) determine the structure of space-time.

The difference between causal precedence and chronological precedence lies in the existence of signals with an invariant velocity c. The chronological precedence can only be established between events a and b if it is possible to send a massive object (e.g. a clock) from a to b. This clock can then measure the

time elapsed between a and b, and hence establish a "chronological" interval between them.

This is not possible for events a and d, connected by a horismotic relation. Here the only possibility for a to influence d, is by using light-like, i.e. massless, signals. For these signals no duration or temporal interval can be determined.

This can be shown with the help of the Lorentz transformations, which state that for an object which is accelerated until its velocity comes close to c durations are stretched out, until time itself almost stops. In the limit where the velocity is c, this means that all clocks have stopped, and that all time intervals have become infinite.

A more physically intuitive argument for this effect is the following: the ultimate way to measure time intervals is by letting a light signal travel back and forth between parallel mirrors which remain at an equal distance (this corresponds to a simplified version of the geodesic clock used by Misner, Thorne and Wheeler, 1974, p. 397). The time between the departure of the light signal at one of the mirrors, and its subsequent arrival after being reflected by the other mirror, is constant, and can be taken as a unit of time duration.

Suppose now that the mirrors are moving in a direction parallel to themselves. Classically, the speed of the reflected light signals should now be equal to the vector sum of their speed when the mirrors are at rest and the speed of the mirrors, hence in absolute value: $\sqrt{2}$ c.

Relativistically, however, the speed of light is invariant, hence equal to c. But since the mirror itself is moving with a speed equal to c, the light signal will not be able to catch up with it, hence it will take an infinite time for the signal to travel from one mirror to the other and back. If the mirrors are moving with a speed somewhat less than c, the light signal will be able to catch up, but it will need more "time" to execute the travel. This explains why this "light-clock" will "tick" slower according as its speed increases, until it stops ticking entirely when its speed reaches c (this last situation can never be attained by an actual, massive clock).

Hence, P establishes a general ordering between events representing the possibility of information transfer, whereas CP establishes an ordering representing the possibility of information transfer whose duration can be established by means of a clock travelling with the signal. The relation HP represents a borderline case of information transfer with the speed of light, and does not establish an order, since the relation is not transitive.

Indeed, if a can be connected to b by a (direct) light signal, and b can be connected to d by a direct light signal, this does not imply that no slower-than-light signal can be used to connect a to d. Consider for example the parallel mirror situation. Call the departure of light at one mirror a. Call its arrival at the other mirror b. Call its arrival back, after reflection, at the first mirror d. Clearly a HP b and b HP d. However a and d correspond to the same spatial position on the mirror, and are only separated by a time interval equal to the period of the clock: hence a CP d, and not a HP d.

On the other hand it is possible to consider P = CP HP, as a transitive closure of HP. Indeed every two events a CP d, separated by a "time-like" interval, can be connected by a series of light signals (e.g. the signals bouncing back and forth in a light clock travelling from a to b).

For a given event a, it is possible to distinguish five sets, which partition the set M of all events:

1) $I^+(a) = \{x \in M: a \text{ CP } x\}$

This is the chronological future of a, which corresponds to the inside of its future light cone through a.

2) $C^+(a) = \{x \in M: a \text{ HP } x\}$

This is the horismotic future of a, corresponding to the boundary of its future light cone.

3) $S(a) = \{x \in M: \text{NOT } (a \text{ P } x \text{ OR } x \text{ P } a)\}$

This is the set of events whose time cannot be compared with the time of a: they come neither before, nor after a. They are separated from a by a "space-like" interval, and correspond to the outside of its light-cone.

4) $I^-(a) = \{x \in M: x \text{ CP } a\}$

This is the chronological past of a, constructed from $I^+(a)$ by inverting the order of the relation.

5) $C^-(a) = \{x \in M: x \text{ HP } a\}$

This is the horismotic past of a.

The relation CP defines a topology on the set M of events: the Alexandrov topology, which is generated by:

$$\{I^+(y): y \in M\} \cup \{I^-(y): y \in M\}$$

(Kronheimer and Penrose, 1967). This is the topology for which a subset of M is open if it is a union of open intervals I (a, b):

$$\forall\, a, b \in M \text{ such that } a \text{ CP } b: I(a,b) = \{x \in M: a \text{ CP } x, x \text{ CP } b\}$$

Such a topology on space-time can be extended to a topology on spatial (three-dimensional) sections of the four-dimensional event set: $S \in M$ is a space-like subset of M *iff*

$\forall$ x, y $\in$ S: NOT ((x P y) OR (y P x)).

This allows us to construct topologies on (state)spaces, whereas in the classical frame the topology is a priori given (cf. section 4.6). Remark that with this topology causal processes are automatically continues, so that it is not necessary to introduce a requirement of continuity (section 4.6).

However, the CP relation is not yet sufficient to construct an unambiguous metric. It can be shown that in general relativity the precedence relation determines 9 out of 10 components of the metric tensor for space-time. This corresponds in special relativity to the fact that the precedence relation determines the Lorentz group (Zeeman, 1964).

If we want to measure distances or durations in an invariant way, however, we still need to assume the existence of rigid rods, determining invariant distances. Once such a distance is determined, it can be used to construct a light clock, and hence it allows us to measure durations.

One way to dispense with rigid rods is provided by the Marzke-Wheeler construction (Misner, Thorne and Wheeler, 1974, p. 397), where a unit of distance or duration is constructed from the paths of freely falling particles (corresponding to the CP relation). The construction, however, is technically rather complex, so that its physical meaning remains obscure. This invariant measurement unit can then be used to establish a metric on space-time. This provides us with invariant metrical relations between events. Invariant means here that the result of the spatio-temporal distance measurement is independent of the specific way of comparison of the distance to be measured with the distance taken as a measuring unit.

We will not go further into the mathematical details of this construction. It suffices here to show that the possible paths of particle-like and light-like signals, which form the causal structure of the event set, are in principle sufficient to reconstruct the basic geometry of space-time, as represented in special and general relativity theory.

Until now, we have only discussed the mathematical properties of the causal relations. We must now also examine their physical and cognitive signification. This will be illustrated by considering the so-called paradoxes which arise if one of the rules defining causal processes is transgressed.

8.4 From local to global causal connections

Although the word "causality" is very often used in relativity theory, mainly to denote the light-cone structure of space-time determined by the P relation, it is never really defined. We know that "causal processes" do not leave the light-cone, but we do not know why. And what must we think of processes which do leave the light-cone, i.e. which travel faster than light? Can they be causal? A serious discussion of this problem is given by Reichenbach (1958).

He considers an example of a process moving faster than light: imagine a searchlight which projects an illuminated spot on a far-away surface (e.g. a cloud). If the searchlight rotates quickly enough, the velocity of the spot will be larger than c. Clearly this movement does not transfer material substances: nothing has been transported from one illuminated part A of the cloud to the next part B, which is illuminated an instant later. The fundamental reason, however, why the spot is able to travel faster than light, is because it does not carry information, hence does not correspond to a causal process. Indeed, someone sitting at A would be unable to manipulate or to "modulate" the beam in such a way that it could transfer a message to someone sitting at B.

In order to explain this in more detail, we must go back to our definition of causality or information transfer in terms of the conservation of distinctions (sections 5.2 and 6.3). This definition is different from Reichenbach's definition (1958), although its application in the present context gives a similar result (for a general overview of different causality concepts, see Apostel, 1975).

Consider two distinctions, one corresponding to the "cause", and the other one to the "effect" of a process. A distinction is characterized by a *marked or indicated state a*, which will be called an *indication,* (cf. Spencer-Brown, 1969), and by the negation or complement of this state a'. Hence, any distinction can be represented by a couple (a, a') of a marked and an unmarked state, or equivalently of a proposition and its negation. (Remark that if we call *a* a state, then we mean "state" in the sense of a representational pattern of activation (section 2.5); on the other hand, in the classical frame, the word "state" is reserved for activation patterns which carry maximum information, i.e. which correspond to atomic propositions (section 4.5); in this last sense a is in general not a state but a proposition.)
We must now define the causality of sets of state transitions constrained by dynamical laws, i.e. of processes.

Definition: a process is *locally causal* if it leads from a to b, and from a' to b', i.e. if it conserves the distinction:

$$(a, a') \rightarrow (b, b') \qquad (*)$$

In other words, all state transitions leading from a to b', or from a' to b are forbidden.

An example of a dynamical constraint, which selects causal processes, is energy conservation. In this case, we have:

a = b = "the energy of the system is E_o", and
a' = b' = "the energy is different from E_o".

A more intuitive example is the following: If I support a stone by keeping it in my hand (a), then it will not fall down to the ground (b). On the other hand, if I

loosen my grip so that the stone is no longer supported (a'), then it will fall down and reach the ground (b').

This definition clearly corresponds to the definition we gave earlier (section 5.2): "equal causes (i.e. situations belonging to the same class a) have equal effects (i.e. situations belonging to the same class b)", "different causes (i.e. situations belonging to a, respectively a') have different effects (i.e. situations belonging to b, respectively b')".

It also corresponds to the possibility of information transmission: if I want to send a signal to someone else, who is spatially separated from me, then I need a physical medium or carrier with at least two states (in the simplest case, the two states correspond to respectively the presence and the absence of the signal). I can prepare the medium in one of both states e.g. a, (i.e. I create an "indication", I mark one of the two states). Through some dynamical mechanism this state will then evolve or propagate until it reaches the other observer to whom the message is addressed. This observer will detect a new state b. In order to understand the meaning of this signal, he must be able to do three operations:

1) to distinguish the state b from its complement b',
2) to deduce from the causal relation ($*$) between the prepared and the detected signal that the original signal I prepared was a,
3) to interpret a, i.e. to determine the extension or meaning of the signal, which is fixed by the language or communication code we have agreed upon to use (this third stage can be seen as repetition of stages 1 and 2, with a now in the role of the signal to be interpreted).

This scheme can be seen as an elementary model of *communication* or *information processing*. It can be used to analyze information transmission as well between two spatially separated subjects, who are communicating, as between a subject and an object he perceives. In the latter case the causal relation and the code correspond to the internal representation of the subject. In both cases we see that an indication, which presupposes a distinction, can really be considered as a "*unit of meaning*" (or "semantic element").

For example, suppose that the purpose of our communication is to synchronize our clocks. The code we have agreed upon is very simple: if my clock indicates 0:00 hours, I will send a light signal towards my correspondent, so that he can adjust his clock to the same time when it arrives. In this case the medium used is the electromagnetic field. The distinction used for communication is that between the state "there is a photon travelling towards the receiver" and the state "there is no photon travelling towards the receiver". The photon itself may undergo a state transition: for example, its frequency may diminish by gravitational red shift. The distinction between photon and no photon, however, remains invariant, so that the received message is unambiguous.

Suppose now that light propagation would not be a causal process. For example, suppose that a photon could disappear, or be created out of nothing. In that case, the message would become ambiguous; it would no longer carry any reliable information. Indeed if the addressee did receive a photon, he would not be able to determine whether it was sent by me, or had appeared spontaneously. If he did not receive a photon, he would be unable to infer whether I did not send a signal, or my signal had disappeared. In both cases, it would be impossible to synchronize our clocks.

This example already illustrates how the determination of space-time frames requires the use of information- or distinction-conserving processes. However, it is possible to go a step further and to show how the structure of space-time can be *derived* from the properties of such processes. In order to do this it suffices to show how the two basic precedence relations (HP and CP) arise naturally out of the properties of information transmitting processes (cf. section 8.3).

The basic characteristic of a precedence relation is its antisymmetry. However, our definition of causal processes is symmetric: the exchange of cause (a) and effect (b), does not alter the relation. This reversibility is a general characteristic of conservation principles (cf. section 5.2). However, it only makes sense for *potential* processes: if a distinction is conserved during a process leading from an event e_1, characterized by a proposition a, to an event e_2, characterized by a proposition b, then it would also be conserved if e_1 was characterized by b, and e_2 by a. In general, however, the actual event e_1 will correspond to a certain state indicated by a, but not by b.

So, we will suppose that actual causal relations between events are generally asymmetric. However, this does not yet exclude the possibility of symmetrical causal connections. Indeed suppose that a signal is sent from $_{e1}$ to e_2, and is sent back from e_2 to e_1. Suppose that each of the two transmissions is causal, i.e. conserves distinctions. Suppose e_1 is determined by a state a. Distinction conservation then implies that e_2 is in a state b, such that:

$$e_1 (a, a') \rightarrow e_2 (b, b').$$

We can use the same reasoning for the reflected signal, and find a correspondence:

$$e_2 (b, b') \rightarrow e_1 (c, c').$$

Hence, there are two possible causal chains: (1) a produces b, and b produces c; (2) a' produces b', and b' produces c'. Suppose e_1 is in a state a, so that process (1) is actual. If a is atomic, we have either $a < c$, or $a < c'$ (section 4.5).

In the first case, the logical relation of implication ($<$) is in correspondence with the causal connection ($\rightarrow$), exemplified by process 1). Hence, the causal connection has not added anything new to the logical-informational structure of the Boolean algebra representing e_1, it has simply confirmed what we already

knew without taking into account the state of the second event e_2. In this sense there was *no information transferred* from e_2 to e_1, no new selection was carried out, everything has remained as it was.

In the second case, the causal connection leads from c' via b, to c. This time, something seems to have changed: after the causal interaction the state of e_1 can no longer be a, since a implies the negation of the new proposition c resulting from process 1). However, the event e_1 is by definition elementary or indivisible in space or time, so that it can have only one actual atomic state. By the law of contradiction, this state cannot imply both c and NOT c. Hence, we come to a *paradox*.

This paradox is equivalent to the traditional causal paradoxes associated with the existence of a "*time machine*", i.e. a device which allows us to influence the past. Suppose that I construct a time machine, and that I use it to go back to the time before my father had met my mother. Suppose that I kill my father. In that case my father will never be able to meet my mother, and hence will be unable to engender me. However, this implies that I shall not be born, and hence that I shall not construct a time machine. We have here the same circular causality as in the reasoning above: the construction of a time machine (c') leads to the killing of my father (b), which leads to the time machine not being constructed (c). This is truly paradoxical.

However, if we remain on the more formal level of distinctions, the contradiction can be eliminated. Indeed if we find that c implies c', and, by applying the causal relation a second time, that c' implies c, we can only conclude that c and c' are equivalent, i.e. indistinguishable. Hence, the distinction between c and c' is logically and physically meaningless, and it should be eliminated from the representation. The conclusion is that the global process which leads from e_1 to e_2 and back to e_1 is not causal, since it does not conserve any real distinction.

The reasoning we made is not only applicable to "causal" paradoxes, but also to "logical" paradoxes. An example of this latter type is the sentence: "the present sentence is false". Suppose the sentence is true (a), then we must believe what it says and conclude that it is false (a'). But if it is false, then we should not believe what it says, and conclude that it is true (a). Hence, we get the same circular structure:

$$a \rightarrow a' \rightarrow a \rightarrow \ldots$$

From the present point of view, which sees representations as based on more or less invariant distinctions, the paradoxical sentence is neither true nor false, since it is impossible to distinguish between both cases. Hence, it does not carry any information or meaning, and does not represent anything at all (except itself).

The same can be said about circular "causal" connections, such as the one mentioned above: there is no conserved distinction at the global level (although the process appears to conserve distinctions on the local level, where only the

one-way connections between e_1 and e_2 are considered), and hence there is no information transferred.

We may conclude that both cases, $a < c$ and $a < c'$, lead to the same result: *no information can be transferred in a circular manner, i.e. from an event to itself.*

The abstract reasoning we made to prove this, can be illustrated with the help of a more concrete, spatial analogy. Suppose that the two sides of a distinction (a and a') are represented by the two sides of a piece of paper (e.g. a corresponds to the upper side, a' to the lower side). A causal process can now be represented by a paper ribbon, which is twisted in a certain way. The left end of the ribbon corresponds to the event e_1 with its two possible states a (upper side) and a' (lower side). The right end corresponds to an event e_2 with two possible states b (twisted upper side) and b' (twisted lower side). The conservation of distinctions is represented by the smooth, continuous shape of the ribbon, which is such that even when it is twisted, both sides can always be distinguished. The movement of a signal from a to b can be represented by a point which travels continuously from the left upper side (a) through the different twists to the right twisted side (b).

If we now want to represent a circular causal process, we must form a closed loop, such that the right end is attached in a continuous manner to the left end. There are two possible ways to do this:

1) between the left end and the right end, which are to be stuck together, the ribbon has been twisted an even number of times,
2) the ribbon has been twisted an uneven number of times.

In the first case, we have something isomorphic to a simple ring without twisting. In the second case we have something isomorphic to a *Moebius band*.

As is well known, a Moebius band is characterized by the fact that it has only one side: if you imagine a paint-brush moving continuously along the paper surface, then it will eventually come back to its initial position after having traveled a distance equal to the double of the length of the ribbon, and having colored the complete surface of the ribbon. It is clearly impossible to make a global distinction between the upper (or outer) and the lower (or inner) side of the Moebius band, since the one is the direct continuation of the other one.

In the case of the simple ring, the two sides can be absolutely distinguished. This means that if a paint-brush begins to paint at one side, it will color this side completely and come back to its starting point without having touched the other side. Hence, if a signal moves around the complete ring, it simply comes back to its starting point, without anything having changed.

8.5 Formal properties of global causal connections.

This analysis shows us that there is a basic difference between relations which are *locally causal* (i.e. conserve distinctions between two directly connected events), and relations which are *globally causal* (i.e. which conserve distinctions between events connected by a *sequence or path* of locally connected events). If the connecting sequence of events is part of a loop or circle, the connection is not globally causal. Only non-circular relations admit global causality. We will now express these properties in a more formal way.

Definition: the relation $C \subset E \times E$ is non-circular, or "*acyclic*" *iff* for any sequence or "path" of events:

$\{e_i \in E: i = 1,\ldots, n \text{ such that } e_j \, C \, e_{j+1} \; (j = 1,\ldots,n-1)\}$

you have that $e_n \, C \, e_1$ implies $e_h = e_k$ for all $h, k \in \{1,\ldots, n\}$

In other words, a relation is acyclic if its (oriented) graph contains no cycles or closed paths.

We can further distinguish two types of global causal connections:

1) either the path connecting two events a and b is unique
2) or there exist at least two paths P_1 and P_2 connecting a to b:

$P_1 = \{e_i \in E \; (i = 1,\ldots n): e_1 = a, e_n = b, e_j \, C \, e_{j+1} \; (j = 1,\ldots n-1)\}$
$P_2 = \{f_i \in E \; (i = 1,\ldots m): f_1 = a, f_2 = b, f_j \, C \, f_{j+1} \; (j = 1,\ldots m-1)\}$

such that there exists at least one $f_i \in P_2$ with $f_i \notin P_1$, or at least one $e_j \in P_1$ with $e_j \notin P_2$.

In other words, P_1 and P_2 pass through different events in their path from a to b.

In the first case we will say that a and b are *horismotically connected*, in the second case their connection will be called *chronological*. It can now be shown that these two types of connection are formally equivalent to respectively the horismotic and chronological relations introduced in section 8.3.

Definition: $\forall$ x, y $\in$ M: x *HP* y *iff* there is a unique causal path P from x to y.

Definition: according to Kronheimer and Penrose (1967), the two requirements for a relation HP to be *horismotic* are:

whenever e_i: i = 1,…n is a finite sequence such that e_i HP e_{i+1} for each i (except n), and h, k are integers satisfying $1 < h < k < n$, then:

(1) e_n HP e_1 implies $e_h = e_k$

(2) e_1 HP e_n implies e_h HP e_k

Theorem: *HP* is a horismotic relation.

Proof: (1) Requirement (1) is equivalent to the requirement that the relation be acyclic. Suppose that P would be part of a cycle, then there would be another path P' connecting b to a such that the concatenation of P and P': P' $*$ P would form a closed path, leading from x, via y, back to x. However, we could then construct a new path from x to y by concatenating the closed path again with P: P $*$ P' $*$ P. Hence, the path connecting a to b would not be unique, in contradiction to our assumption.

(2) Suppose that $_{e1}$ *HP* e_n, i.e. there is a unique path $P_{1,n}$ from e_1 to e_n. We know that e_i R e_{i+1}, i.e. the path $P_{i,\,i+1}$ between e_i and e_{i+1} is unique. The concatenation:

$P_{n-1,n} * P_{n-2,n-1} * \ldots * P_{1,2}$

defines a path from $_{e1}$ to e_n, which by assumption is unique and equal to P_{1n}. The segments of this path determine the unique paths connecting e_h to e_k ($1 < h < k < n$):

$P_{hk} = P_{k-1,k} * P_{k-2,k-1} * \ldots * P_{h,\,h+1}$.

Hence: e_h *HP* e_k ■

Definition: x *P* y *iff* there is a causal path from x to y, which does not form part of a cycle.

Definition: x *CP* y *iff* there are at least two paths, which are not part of a cycle, connecting x to y.

Property: clearly we have: $P = CP \cup HP$ (disjoint union)

Definition: the quadruple (M, P, CP, HP) where P, CP and HP are relations on M, is a "*causal space*" *iff* the following conditions are satisfied (cf. Kronheimer and Penrose, 1967): for x, y, z $\in$ M:

(1) x P x
(2) if x P y and y P z then x P z
(3) if x P y and y P x then x = y
(4) not x CP x
(5) if x CP y then x P y
(6) if x P y and y CP z then x CP z

if x CP y and y P z then x CP z
(7) x HP y if and only if x P y and not x CP y

Theorem: the relations *P*, *CP* and *HP* on M as defined above, form a causal space.

Proof:
(1) x is connected to x by the trivial, causal path (exemplified by the distinction conserving relation of identity)
(2) if there is an acyclic path from x to y, and an acyclic path from y to z, then the concatenation of both paths forms an acyclic path connecting x to z.
(3) suppose $x \neq y$, then the condition x *P* y and y *P* x would imply that there is a non-trivial path from x, via y, back to x, and this is in contradiction with the assumption of acyclicity.
(4) if there would be two different paths connecting x to itself, then at least one of these paths should be non-trivial, and would hence correspond to a closed loop.
(5) this requirement follows directly from the fact that *CP* is included in *P*.
(6) x *P* y signifies that there is at least one path from x to y, y *CP* z signifies that there are at least two paths from y to z. By concatenation we find at least two different paths leading from x to z.
(7) this condition follows directly from the fact that *P* is a disjoint union of *CP* and *HP*. ■

Conclusion:
Suppose that you have a set M of potential events, and a locally causal relation C $M \times M$, defining direct causal connections between events, then this relation can be extended to include indirect causal connections, i.e. paths or sequences of direct causal connections. There are four different types of possible indirect connections between two events x and y: suppose

$$P = e_1,\ldots, e_n, \text{ with } e_1 = x, e_n = y, e_j \; C \; e_{j+1}, j = 1,\ldots n$$

is a causal path connecting x to y, then there are two possibilities:

1) P is part of a cycle, i.e. there exists another path P' from y to x.
2) P is not part of a cycle.

Both categories can be further subdivided in two subcategories:

1.1) P' $*$ P conserves distinctions globally, i.e. the indication transferred from x to y and back to x is not in contradiction with the initial indication determining the state of x. In this case there is no real transmission of information, but there is a *correlation* between the state of x and the state of y.

1.2) P' $*$ P does not conserve distinctions on a global level. In this case we would find a causal paradox, if we would interpret the transfer of distinctions

literally. The only way to avoid this, is by regarding the supposedly transferred distinction as void or meaningless. Hence, this case is equivalent to the case where there are *no causal connections at all* between x and y.

2.1) The path P is unique. This also entails that all segments $\{e_h, \ldots, e_k: 1 < h < k < n\} \subset P$ are unique. In that case the signal travelling along P can be interpreted as a light signal, i.e. a signal with an invariant limit speed, moving in a straight line. The connection is *horismotic*.

2.2) There is at least one different path P' leading from x to y. P can be interpreted as a *chronological* connection. This was shown formally. We will now try to explain this result on a more intuitive, physical level.

The characteristic of a chronological connection is that it is always possible to have a faster connection: suppose that x corresponds to a clock A indicating a time t_A, and y to a clock B (in general spatially separated from A) indicating a time t_B. The chronological connection between x and y signifies that there is a chain of subsequent signals of which the first one departs from A at time t_A, and the last one arrives at B at time t_B.

If another, faster sequence of signals would depart from A at time t_A, this would reach B at a time $t'_B < t_B$. The concatenation of this sequence with the causal process consisting of the movement of the clock B between t'_B and t_B, would form an alternative causal path connecting x to y. Hence, the path would not be unique. On the other hand, a horismotic or light-like path connecting two events is necessarily unique.

We must further explain why horismotic signals have an invariant speed. In order to define a speed, you need to establish a duration, and this requires either a clock which travels along with the signal, or a set of two synchronized clocks, one at the arrival and one at the departure of the signal. However, the fact that the causal path of a horismotic signal is unique, implies that no clock can travel in parallel with it, since this would create an additional sequence of causally related events (e.g. the subsequent "ticks" of the clock). On the other hand, the only way to synchronize two spatially separated clocks in a more or less invariant way, is by using light signals. Hence, we need to assume an invariant speed for horismotic signals in order to determine their speed!
This is another formulation of the general principle of the *impossibility of measuring the one-way velocity of light* (Podlaha, 1980). The only thing we can measure in an invariant way is the two-way velocity, by determining the interval between the departure of a light signal and the arrival of its reflection. A reflected light signal, however, does not correspond to a horismotic connection, but to a chronological one (cf. section 8.3), since it allows a parallel causal path for the clock measuring the interval.

This clock can be assumed to be a light clock, or, more generally a geodesic clock. The reflecting light signal itself can be viewed as forming another, larger light clock moving in parallel with the first one. The invariance of the two-way speed of light then implies that if both clocks were once synchronized, they remain synchronized. In other words, both causal paths formed by the two light signals moving back and forth between reflecting mirrors, remain in phase; it is

not so that one clock will begin to tick faster with respect to the other one (cf. Misner, Thorne and Wheeler, 1974, p. 398).

This is an empirical fact which can as yet not be explained from the structural properties of distinction conserving processes. However, once this fact is assumed, it becomes a matter of convention to determine an invariant one-way speed for horismotic signals.

On the other hand, the topological and causal structures of space-time follow directly from our analysis of distinction conserving relations (cf. section 8.3).

The interdependence of topology or continuity and distinction conservation can be demonstrated in an even more direct way, without passing through the intermediate stage determined by the precedence relations. Indeed, continuity is defined mathematically as the conservation of topological structures. In other words, an operator is said to be continuous if it maps closed sets onto closed sets, and open sets onto open sets. But a closed set can be viewed as the union of an open set and its boundary. Hence, continuous transformations are transformations which conserve boundaries. But a boundary is just that which separates or distinguishes the inside of a spatial domain from its outside. Hence, continuity is equivalent to the conservation of spatial distinctions, and the topological structure of a state space is just a representation of certain distinctions which are conserved by all causal processes on the system.

8.6 Non-locality paradoxes in quantum mechanics

The reasoning we made to demonstrate the dependence of topology, and hence locality, on information transfer, can be used to elucidate some paradoxes where the interaction appears to be non-local. These paradoxes are typical for quantum mechanics. This should not surprise us, since the characteristic feature of quantum mechanics is that distinctions are in general not conserved. We will consider three such paradoxes: the paradox of de Broglie (de Broglie, 1959), the Aharonov-Bohm effect (Aharonov and Bohm, 1959) and the EPR paradox (Einstein, Podolsky and Rosen, 1935).

8.6.1 The paradox of de Broglie

The de Broglie situation is formally and conceptually the simplest and can be considered as a prototype for all apparently non-local effects in quantum theory. Consider a quantum system imprisoned in a box, so that its wave function is continuously spread out over the volume of the box. The box is then divided in two by the introduction of an impenetrable plate. The two compartments A and B are then separated by an arbitrarily large distance. Normally each compartment will contain about half of the quantum wave. Suppose we now make a measurement to test whether the system is present in box A. Suppose the result is negative.

According to the postulate of the reduction of the wave packet, the part of the wave function localized in box A will have vanished after this operation. The total probability for finding the system in one of the two boxes however,

has to remain constant. This implies that the probability of finding the system in the second box B, and hence the absolute value of the wave function in B, will become larger.

It appears as though a part of the function has been transferred instantaneously from A to B. Yet we have postulated that A and B are separated by an impenetrable barrier, and by an arbitrarily large distance. Hence, this strange jump of the wave function from A to B seems to be in direct contradiction with the principle of locality which states that every physical interaction between two spatial domains A and B must be carried by some physical medium connecting A to B, and hence cannot travel faster than the speed of light.

However, as seen from the viewpoint developed in the previous section, there is no contradiction. The locality condition is really a requirement of continuity for physical processes, which prohibits discontinuous jumps over topological boundaries. As we have shown, this requirement only applies to information transferring processes, i.e. acyclic, distinction conserving sequences of events. It is shown easily that the collapse of the wave function, from box A to box B, cannot be used to transfer information.

The basic distinctions in the de Broglie situation are those between the presence and the absence of the system in the box A (respectively B). Let us call these distinctions (a, a'), respectively (b, b'). There is clearly a correlation between both distinctions: the indication a implies the indication b' (if the system is present in A, then it must be absent in B), and a' implies b. So, there is a conservation of distinction:

$$(a, a') \leftrightarrow (b, b')$$

However, the relation between both distinctions is symmetric, since b implies a', and b' implies a. The relation thus falls in the category 1.1) of cyclic, globally distinction conserving connections. Such relation, however, cannot be used to transmit signals, since it does not produce an observable (i.e. distinguishable) change.

Indeed, to transmit information from A to B we must prepare the system at A in a certain state (e.g. a') so that our correspondent, located at B, can detect the corresponding, causally connected state (e.g. b). The problem is that we cannot choose to prepare the system so that a' is necessarily actual. According to the assumption that the original box containing a homogeneously diffused wave function, has been split in two equal parts, there is a fifty percent chance to find a, respectively a'. After the separation has been carried out, we have no way to make sure that we will find the state a' when opening the box.

The only thing we can do is to observe which of both alternatives, a or a', is actual. If we find that a' is actual, then we know that our correspondent at B will find b. However, our correspondent does not know what he will find. Whether we have observed the system at A, and hence reduced the wave packet, or not, for him the probability of finding b, respectively b', remains fifty percent.

Hence, he cannot infer anything about the way we have manipulated the system at A by looking at B. No information has been transmitted.

This result may appear trivial when it is compared to a seemingly analogous classical thought experiment. Suppose that the original box does not contain a quantum particle, but a classical system, e.g. a billiard ball. Suppose that we do not know the exact position of the ball in the box. If the box is now again split in equal compartments, then we can assume that there is a fifty percent probability to find the ball in one of the compartments, say A. Finding the ball in A signifies that another observer opening box B will find it empty. Again there is a symmetrical conservation, i.e. a correlation, of distinctions between both boxes.

However, this can be explained simply by noticing that the two observations at A and at B had a *common cause,* namely the original position of the ball in the box. If the ball was somewhat left of the middle (l), it would become enclosed in the left compartment (e.g. A) when the dividing plate was introduced, otherwise (l') it would be enclosed in the right box (B). This can be represented by two causal connections:

(l, l') ↔ (a, a') and (l, l') ↔ (b', b)

The original distinction between left and right has simply determined the distinctions between presence and absence in box A, respectively B. The distinction conserving relation "has the same cause as" between (a, a') and (b, b') is clearly symmetrical, and hence accounts for their correlation.

We could now wonder whether an analogous explanation could be found for the quantum correlations. Such an explanation would correspond to a hidden variable model, i.e. a classical statistical model, based on the assumption that the quantum particle is already located in one of the two parts, left or right, before the compartments are separated. This is in contradiction with the traditional interpretation of the wave function, i.e. of the quantum representation. However, in the present, simple situation it would provide an equivalent model of the experiment.

8.6.2 The EPR paradox

This is no longer true for the more complex situation exemplified by the EPR paradox. Suppose that a quantum system, with a total spin equal to zero, disintegrates in two particles of the same type, e.g. photons, each with a spin equal to one half. By the law of spin conservation, the directions of the two spins must be opposite, so that their sum remains zero.

This means that if we make on observation on one of the particles, and find that its spin is up, than the wave function of the other particle must collapse so that its spin is necessarily down. This instantaneous effect is again independent of the distance, or of eventual barriers between the two particles. This is completely analogous to the de Broglie situation.

The difference, however, is that we can choose between different observations of the particles, and that each time we will find the same

correlation. For example, we could decide to measure the spin in the left-right direction, instead of the up-down direction. Again we will find that if particle A has spin left, then particle B has spin right. This cannot be explained by a classical common cause for the two results, because this would require that the particle A has an initial state where the spin is at the same time left and up. But the properties of spin up and spin left cannot both be actual.

A mathematical proof of this impossibility of explaining the quantum correlations by assuming hidden variables for the two particles was given by Bell (1966). He considers *coincidence experiments* where two simultaneous yes-no measurements a and b are made, one on particle A and one on particle B. Let us define a variable describing the possible results of such an experiment (the present notation corresponds to that of Aerts, 1982):

$X_{ab}(s) = +1$ *iff* for a system in the state s, we get the result "yes" for particle A, and "yes" for particle B, or "no" for A, and "no" for B

$X_{ab}(s) = -1$ *iff* we get the result "yes" for A, "no" for B, or "no" for A and "yes" for B.

Now define a variable describing the possible results of a single experiment, on only one of the particles:

$X_a(s) = +1$ *iff* $s < a$

$X_a(s) = -1$ *iff* $s < a'$

Bell now assumes that there is a classical *locality condition*, and concludes that necessarily:

$$X_{ab}(s) = X_a(s).\ X_b(s) \qquad (1)$$

In other words, the variable describing the coincidence experiment can be factorized in two variables describing the two single experiments on the different particles. This signifies that the result of the measurement a on A does not change the result of the measurement b on B. Otherwise we could have that a state s, which would give the result "yes" for a, and "yes" for b, if a and b were performed separately, would give a result "yes" for a, "no" for b if a and b were performed together. In that case we would have:

$X_a(s) = +1$, $X_b(s) = +1$ and $X_{ab}(s) = -1$,

contrary to Bell's locality condition.

At first sight, Bell's assumption appears perfectly natural: if the initial state s is such that both a and b have determined answers, when performed separately, and if the two measurements are performed together, but so that they are

separated by a space-like interval, so that no signal can be transmitted from one measurement event to the other one, then we should find the same determined answers for the coincidence experiment.

Of course, in the quantum situation, the quantum state generally does not allow us to determine the answers for the measurements a and b. However, Bell assumes that this is due to a classical lack of knowledge, and that there exists a classical microstate, which we do not know, but which determines the answers.

Bell then uses his assumption (1) to derive the following inequality:

$$X_{ab}(s) - X_{ac}(s) + X_{db}(s) + X_{ac}(s) < 2,$$

where a, c are measurements on A, and b, c are measurements on B.

(this result follows trivially by factoring X_a (s) in the first term, and X_d (s) in the second term)

However, in the quantum correlation experiment, the Bell inequality is generally violated. This has been shown as well theoretically, by using the quantum formalism, as empirically (Aspect, Dalibard and Roger, 1982). Therefore, we must conclude that Bell's locality assumption does not apply for quantum systems. Hence, we cannot explain quantum correlations between spatially separated systems by assuming a "common cause" s for the correlated events, which determines independently the results of the experiment at A, and the experiment at B.

These mysterious correlations are not restricted to the microscopic world, however. Aerts (1982) has proposed a macroscopic system which violates the Bell inequality. The system consists of two vessels A and B (standing on the same horizontal plane) connected by a tube of arbitrary length. The system is filled with water, and the measurements under consideration consist of emptying one of the vessels by means of a siphon, and checking whether the amount of water collected thus in a reference vessel is more than 10 liter.

Let us call such a yes-no observation a when it is performed on vessel A, and b when performed on vessel B. Suppose each vessel contains 10 liter. This corresponds to the state s of the system. If we empty vessel A with a siphon, we shall clearly get more than 10 liter. Indeed, when the water level in A goes down, water from B will flow through the tube to compensate, so that the total amount of water which can be collected in this way will be near to 20 liter. Hence, observation a will always produce the result "yes" in state s. The same is true for b.

However, if we perform both measurements at the same time, the total amount of 20 liter will have to be distributed in some way between the two reference vessels at A and at B. Since normally the flow of water through the two siphons will not be completely equal, we may expect that the water level in one of the reference vessels (say at A) will rise a little faster than the level in the other vessel. When the water level in the two connected vessels reaches the bottom, the water flow will stop in both reference vessels at the same time, so

that the reference vessel at A contains somewhat more water than the other one. Since the total amount to be collected is 20 liter, after the experiment the reference vessel near A will contain more than 10 liter, whereas the reference vessel near B will contain less than 10 liter.

Hence, when performed together, the two observations a, respectively b, produce the result "yes", respectively "no". Hence, we get:

$$X_{ab}(s) = -1 \text{ whereas } X_a(s) = +1 \text{ and } X_b(s) = +1.$$

This is in contradiction with Bell's locality assumption (1).

Let us analyze this thought-experiment by means of the distinctions framework. The observation a corresponds to the distinction:

(a, a') = (**the amount of water in vessel A is more than 10 liter, the amount of water in vessel A is less than 10 liter**).

Analogously, the observation b corresponds to the distinction (b, b'), where vessel A has been exchanged for vessel B. The coincidence experiment of a and b shows a correlation between both distinctions:

(a, a') ↔ (b', b)

However, this correlation cannot be explained by a third distinction acting as a common cause, like in the example of the billiard ball in the box. This can be illustrated by considering the situation where the tube connecting the two vessels A and B is closed. In that case, it is possible to ascertain which vessel contains the largest amount of water by doing one observation, e.g. a. The water collected in this way comes only from vessel A. If the total amount of water is 20 liter, then we know that if A contains e.g. 12 liter, than B must contain 8 liter, so that observation a will produce the result "yes", and b "no". The coincidence experiment will produce "yes, no". Hence, the Bell locality condition is satisfied, and the distinction:

(**the largest amount of water is contained in A,**
the largest amount is contained in B)

can be considered as a common cause for (a, a') and (b, b'). This is completely analogous to the billiard ball situation, where the initial distinction is:

(**the center of gravity of the ball is in compartment A,**
the center of gravity is in compartment B).

However if the connecting tube is open, it becomes impossible to determine which vessel contained initially the largest water volume, by doing a single experiment, e.g. a.

Moreover, if we would be able to determine the volume in another, indirect way (e.g. by measuring the dimensions of the vessel and calculating its volume), this would not allow us to predict the result of the coincidence experiment. Indeed, the volume collected from vessel A does not depend on A's volume, but on the speed with which the water flows through the siphon. The only distinction which does influence the result of the coincidence experiment is that between the sizes of the two siphons:

(**the siphon at A is larger, the siphon at B is larger**) > (a, a').

However this distinction does not belong to the system which is observed, but to the measuring apparatus. Hence, it corresponds to what we have called the indeterminate microstate of the measuring instrument (sect 7.3). At the level of the system to be observed there is no initial distinction:

(**there is more water to be collected from A,**
there is more water to be collected from B)

as long as the tube is open. The distinction is really *created* during the observation process, and depends upon the observation apparatus. This is completely in correspondence with our analysis showing that the quantum observation process does not conserve distinctions.

As we have continuously emphasized, processes which do not conserve distinctions, cannot transfer information, and hence are not subject to the requirements of causality, locality or continuity. It can indeed be proven in a rigorous way that the EPR set up cannot be used to transmit signals (Ghirardi, Rimini, and Weber, 1980). Hence, there is no paradox in the fact that the EPR correlations are non-local, i.e. are correlations between events separated by a space-like interval.

The apparent paradoxical nature of the EPR effect is simply due to the fact that the correlated, spatially separate distinctions are not caused by a common initial distinction, corresponding to a classical, local hidden variable, as shown by the Bell theorem. However, there is nothing mysterious about that, if the phenomenon is analyzed in a more general, non-classical frame, where there is room for the creation of distinctions by the observer, and for a symmetrical, distinction conserving relations between events. In such a frame the Bell locality condition (and hence the Bell inequality) would generally not be valid, because it presupposes that the distinction made during a coincidence experiment was already inherent in the state of the object, and hence could be uncovered by single experiments. The Aerts thought experiment has shown that this is not even true for macroscopic systems.

8.6.3 The Aharonov - Bohm effect

We will now apply our analysis to a last quantum mechanical paradox, which is at first sight of a different nature: the Aharonov-Bohm effect. The non-local

correlations appear here to be independent of the observation process represented by the collapse of the wave function; they depend only on the dynamical part of the quantum representation, characterized by the Schrödinger equation.

Consider a magnetic field confined to an infinitely extended solenoid. Consider an electron wave function confined to the spatial domain outside the solenoid. Clearly there is no local contact between field and wave. Yet the solution of the Schrödinger equation for the wave function can be shown to depend on a parameter which measures the flux of the magnetic field through the solenoid.

The reason is that the Schrödinger equation depends on the magnetic potential, which is non-zero outside the solenoid, and which thus gets into the expression for the phase factor of the wave function. This flux dependence of the phase factor can be observed by doing an interference experiment with the wave. The interference pattern will undergo a shift proportional to the flux.

This effect can again be represented very simply by a correlation between two distinctions: call the situation where there is a shift in the interference pattern s, and the situation where there is a non-zero magnetic field in the solenoid m. We then have a correlation:

(s, s') ↔ (m, m').

This correlation is again non-local, since the events of detecting whether s or s' is the case respectively detecting whether m or m' is the case, can always be produced by a coincidence experiment, so that they are separated by a space-like interval.

Let us analyze whether this correlation can be used to transmit signals. There are two possible cases: either the magnetic field has always been present in the solenoid, with a constant flux, or at some moment in time the magnetic field has been turned on.

In the first case, the correlation cannot transfer information, just like the de Broglie, or EPR correlations. Indeed, the state of the distinction (m, m') is fixed, and hence, through the correlation, the state of (s, s') is fixed. Therefore, we cannot prepare the field or the electrons in a particular state, we can only detect what is already the case, and predict what our correspondent looking at the other distinction will find. However, this correspondent has no way to establish whether we have looked at the distinction or not.

In the second case, we can prepare the magnetic field in a certain state, e.g. by turning it on if it was turned off. This will influence the state of the electron wave, so that our correspondent looking at the interference pattern (which can be at a large distance of the solenoid), can detect the pattern shift caused by this event. Hence, a signal has been transferred.

However, according to the laws of electrodynamics, if the magnetic field is turned on, this must necessarily be accompanied by the appearance of an electric field, extending outside the solenoid. This electric field will interact locally with

the electron wave by shifting its phase. When the magnetic field becomes constant again, the electric field will disappear, leaving behind a shifted electron wave. This local effect of the electric field on the wave, is subject to the laws governing electromagnetic propagation, and hence must spread with a velocity which is equal to the velocity of light. Hence, the velocity of the transmitted signal cannot be larger than the velocity of light (cf. Van Kampen,1985). Therefore there is no paradox.

Remark that in this second case the correlation can be explained by a common cause. Indeed the turning on or off of the field has directly caused (m, m'), and indirectly, through the appearance of the electric field, caused (s, s').

In the first case, where the field is absolutely constant in time (this is clearly not a realistic situation), the origin of the correlation can be traced back to the fact that the Schrödinger dynamics conserves superpositions of quantum states. Indeed, the changing of the phase factor can only be made observable because waves travelling along opposite sides (left or right) of the solenoid acquire different phase factors, and hence will produce a shifted interference pattern. Classically, an electron can only travel along one path, left or right of the solenoid, so that there is no interference of different phases.

This can be represented by a *topological distinction*: the space outside the solenoid has a multiply connected topology. This signifies that paths travelling along different sides of the solenoid cannot be continuously deformed one into the other; they belong to different "homotopy classes ". The solenoid can be viewed as a "hole" in space forming an absolute boundary separating paths which belong to different classes.

However, quantum observation processes do not conserve topological distinctions. This means that a general quantum state will be a superposition of waves travelling along different homotopy classes. Such a superposition state does not distinguish between homotopy classes. This is analogous to the superposition state in the de Broglie situation which does not distinguish between the separate component spaces (boxes) of a non-connected space, and to the state of the connected two-vessel system in the Aerts thought experiment, where the water is distributed in such a way that it is impossible to distinguish between the vessel containing the larger amount of water and the vessel containing the lesser amount.

If we wish to make distinctions between different paths, we must make the wave function collapse to one of its position eigenstates. This is what happens during the interference experiment when the wave is projected onto the screen. The interference pattern is really a two dimensional image of the correlations between different measurement results (i.e. points on the screen hit by different particles). The fact that this pattern itself depends on the magnetic flux is then analogous to the fact that the spin correlations of the EPR experiment depend upon the initial spin of the particle which has disintegrated.

CHAPTER 9: Irreversible information processes: from statistical mechanics to cognitive psychology

9.1 Introduction

We have seen that the basic difference between classical and non-classical representations is to be found at the level of the information processes within the representation, which simulate and hence allow us to anticipate external processes. In classical representations these feedforward processes conserve all distinctions, hence conserve the amount of information. Non-classical representations are characterized by internal processes which create or delete distinctions: e.g. the observation process in quantum mechanics, or the "paradoxical" cyclic causal processes in space-time physics.

In these examples, however, the non-conservation of distinctions seemed to have rather negative effects. Each time this phenomenon appears in a theory, it gives rise to indeterminacy or relativity. Furthermore, it is accompanied by a host of paradoxes, confusions, and ambiguities. This is because the classical frame, which unconsciously determines the world view of most scientists, is unable to incorporate these phenomena.

In the examples we shall discuss now, however, the creation and destruction of distinctions will be seen in a more positive way. We shall show that this phenomenon is a very natural one, which is needed to explain all evolutions where something qualitatively new is produced. Due to lack of space and time, however, the discussions must remain very superficial.

9.2 Irreversibility in statistical mechanics

As we have said earlier (section 7.4), the main difference between classical mechanics and statistical mechanics lies in the fact that statistical mechanics uses macrostates to represent the *incomplete information* an observer has got about a *complex system*.

A typical example of such a system is a box filled with a gas. The number of gas molecules in the box is so great, that it is impossible to determine the basic state variables (position and momentum) of all the individual molecules. The only things the observer can determine are global, macroscopic variables (volume, pressure, temperature,…), which describe the basic properties of the gas as a whole. The values of the macroscopic parameters determine the macrostate of the system, which can be viewed as the subset (or ensemble) of all these microstates, which imply the given values. (More generally, a macrostate can be viewed as a probability distribution on the space of the microstates.)

One would then be tempted to represent the evolution of the complex system by an ensemble of microstate trajectories determined by the classical, deterministic laws of microstate evolution. If the macrostates are well-chosen, so that the set of microstates determining a macrostate m(t) is mapped onto a

new set which corresponds to another macrostate m(t+T), then the evolution of macrostates within the macrostate space will be deterministic.

In that case, the transition from microstates to macrostates amounts simply to a "reduction of the representation" (cf. section 4.8), where the new representation is homomorphic to the old one. The new representation is obtained from the old one by eliminating all distinctions which are considered irrelevant, i.e. those distinctions which discriminate between different microstates belonging to the same macrostate. Although the new representation will contain less information than the old one, it will still be classical, i.e. it will conserve all remaining distinctions.

It can indeed be shown that the classical, reversible evolution conserves the volume of regions of state space (Liouville theorem). This volume can be seen as a measure for the amount of microstates belonging to a macrostate (assuming that the probability density of microstates is constant within the volume, and zero outside). Hence, the "number" of microstates belonging to a macrostate is conserved, and this means that the number of distinctions needed to determine the microstate remains constant.

This property can be clarified by introducing the concept of *entropy* (S), which is an additive measure for this number of distinctions (i.e. the amount of information) needed to determine a microstate s, given a macrostate m:

$$S(m) = k \log W(m)$$

where k is a constant (often the constant of Boltzmann) and W the number of microstates belonging to m. This formula was introduced by Boltzmann. It can be generalized for situations where the probability distribution is not constant:

$$S(m) = - \sum_i P_i(m).\log P_i(m)$$

where $P_i(m)$ is the probability to find a microstate s_i provided that m is actual. (In the continuous case, the sum will be replaced by an integral).

This is the classical expression for the amount of lacking information. If k is chosen equal to 1, and if the logarithm has base 2, S corresponds to the "*number of bits*", i.e.: *the number of (equiprobable and independent) distinctions to be carried out in order to determine s, given m*.

We have not used this measure until now, because in general the two sides a and a' of a distinction are not equiprobable, and different distinctions are generally dependent (e.g. a.b would have a greater probability than a'.b). However, in general there is no absolute way to determine these probabilities.

In the framework of statistical mechanics, were all variables are supposed to have determined probabilities, this measure is very useful. It is mostly used to express the *reversibility* (or *irreversibility*) of a specific evolution. Indeed, according to the classical, causal dynamics, all information is conserved, and hence S is *constant for reversible processes*.

However, it is easy to imagine situations where the incomplete, macroscopic information of the observer will diminish. A traditional example is the following: suppose that you have a box, separated by a plate in two compartments A and B. Box A contains a gas, box B is empty. If the plate is removed, the gas molecules will diffuse into B. After a while, the distribution of the gas over the two compartments will have become homogeneous. Call (a(x), a'(x)) the distinction:

(**the gas molecule x is in box A, the gas molecule x is in box B**)

Before the removal of the plate, it was possible to determine which of both alternatives was actual (namely a), for all the molecules x of the system. However, after the removal, we can no longer ascertain whether an individual molecule is in box A or in box B. The same thing would have happened if all the molecules were originally present in box B.

Hence, we have lost the information about which compartment contains the molecules. The distinction (a, a') has become unobservable, hence meaningless. This signifies that the new macrostate m_1, representing the diffused system, will contain less information than the original macrostate m_o, representing the system contained in A. Hence:

$$S(m_1) > S(m_o).$$

In other words, m_1 contains more microstates than m_o. If we suppose that all microstates are equiprobable, this means that m_1 is more probable than m_2.

Let us estimate how much m_1 is more probable than m_2. Suppose that the number of molecules in the system is N (let us say equal to the Avogadro number, i.e. very large). Each molecule can be in two distinctional states: a or a', assuming that m_1 is given. Hence, there are 2^N subsets of m_1 to be distinguished if we could determine whether a molecule was in A, or in B. One of those subsets, where all molecules are in A, corresponds to m_o. Hence, m_1 is 2^N times more probable than m_o.

In other words, the chance of finding m_o, given the global constraints expressed by m_1, can be neglected. The macroscopic or global properties determining the system, such as global energy, volume, number of particles,…, allow different configurations of the systems elements. However, the configuration represented by m_o, where all molecules are concentrated in one half (A) of the box, is so improbable, that we may assume that we will never see it occur spontaneously. Hence, we can assume that the configuration to be expected is the one where all molecules are homogeneously scattered over the entire box, i.e. the macrostate which does not distinguish between compartments A and B.

If this idea is formulated as a dynamical constraint, we get the *second law of thermodynamics*:

a closed system will always evolve towards its most probable macrostate, i.e. the macrostate m for which the entropy S (m) is a maximum.

This principle is generally applicable to all complex or thermodynamic systems, i.e. systems with a very large number of components. For systems with a small number of components we cannot neglect the probability of macrostates for which the entropy is not maximal, hence the principle is not universally applicable.

The main consequence of this principle is that thermodynamic evolution becomes *irreversible*: a thermodynamic evolution towards equilibrium (i.e. maximum entropy) cannot be reversed, because this would mean that entropy would diminish. The equilibrium state acts as an attractor.

This irreversibility contradicts classical distinction conservation, because distinct initial states e.g. (all molecules in A, all molecules in B), can evolve towards the same equilibrium state. This confronts us with a paradox, because we started with the assumption that statistical mechanics is based on classical, reversible dynamics.

If we wish to express the irreversible thermodynamic evolution as a deterministic trajectory in the space of distributions over classical microstate space, then we clearly cannot use the classical dynamical constraints exemplified by the Liouville theorem. We must introduce *stochastic processes*, i.e. processes where for a given initial microstate there are different potential transitions, each with its proper probability.

If these processes are not time symmetric, i.e. the transition probability P (s_f s_i) is in general different from P (s_i s_f), it is possible to define a function S of the probability distributions, whose time derivative dS/dt is positive. This function corresponds to the entropy, and is a measure of the irreversibility of evolution.

In other words, to represent irreversible thermodynamic evolutions, it does not suffice to introduce a lack of information about the state, because this would lead to a model with non-zero, but constant entropy. We must add a lack of information about the dynamical constraints, i.e. we must allow different possible state-trajectories starting at the same initial state.

A good exposition of this tension between reversible and irreversible representations in statistical mechanics is given in (Prigogine, 1979). However, Prigogine's attempt to solve the paradox appears rather artificial. He tries to define a transformation mapping distributions from the reversible representation onto distributions from the irreversible representation. This would also transform the operator group representing the reversible evolution into an operator semigroup (i.e. no inverse operators) representing the irreversible evolution. The physical meaning of this transformation, however, remains vague.

From the present point of view, there is no reason to derive the irreversible representation from the reversible representation. If we are willing to accept that it is impossible to get complete information about the *state* of a thermodynamic system, then we should also be willing to accept that it is impossible to get complete information about the system's *dynamical constraints*, since these depend upon the unknown state variables.

For example, if we do not know the momentum of two molecules in a gas, then it is meaningless to use the principle of momentum conservation in order to determine the directions in which they will be scattered. Even if we had (incomplete) information about their initial positions (e.g. inside compartment A), then we would have lost this information after a few unpredictable interactions.

The only constraint we can use is continuity: a molecule cannot move instantaneously from the center of compartment A to the center of compartment B. Apart from that, and some global constraints (conservation of total energy, total momentum, etc.), there is no way to restrict the number of possible state transitions. Hence, the information we had about the initial state can only diminish during further evolution, until equilibrium is reached. In other words, the distinctions we were initially able to make are subsequently lost.

We find here an analogy with quantum observation processes. Suppose a quantum system is prepared in a state with definite momentum. Then its position is measured by a macroscopic apparatus. As we have seen (section 7.3), the interaction depending upon the unknown microstate of the apparatus will perturb the state of the quantum system in an unknown way, and hence destroy the information about the momentum of the particle. Here too the lack of information about the microstate (of the apparatus) produces a lack of information about the interaction, and hence about the further evolution of the coupled system (apparatus + quantum object). However, this loss of information about momentum is here compensated by a gain of information about position.

We shall now see that in certain cases, a thermodynamic system can also experience a gain in information.

9.3 Self-organization

The second law of thermodynamics states that entropy cannot diminish, hence information cannot increase, in closed systems. A closed system is a system which does not interact (i.e. does not exchange information) with its environment. However, if the system is open, then it is possible that the entropy produced in the system is evacuated or *dissipated*, so that the information inside the system increases. This gives rise to new distinctions between parts of the system.

For example, a layer of water heated from below, can give rise to the *Bénard instability*: a structure of hexagonal cells appears in the liquid. Inside the cells, warm water moves up from the bottom to the relatively cool surface. Outside the cells the water which was cooled by its contact with the surface

sinks downward again to the bottom. Before the water was heated there was no distinction between inside and outside of a cell; the movement of water molecules was homogeneously spread. The heating, however, has created a non-equilibrium, resulting in a dissipation of entropy. This becomes visible as a structure of distinct cells.

Such phenomena were called "*dissipative structures*" by Prigogine (1979). Analogous phenomena are sometimes designated by "spontaneous symmetry breaking", "order from fluctuations", or "synergetic systems" (cf. Haken, 1978).

This terminology can easily lead to confusion, since at first sight we would tend to associate order with more symmetry, not with less symmetry. On the other hand, the concepts of "dissipation" and "synergy" are more readily associated with the classical concept of "energy", whereas the phenomena we try to describe are better characterized by "information". Their essential characteristic is the appearance of *structure*, i.e. internal differentiation and integration, or, in our terminology, distinction and assimilation.

Since these phenomena are in general dynamic, it is better to replace the word "structure", which has a static connotation by the word "organization". Since the phenomenon appears spontaneously, inside the system, we will speak about "*self-organization*": there is no external agent determining the organization.

From a purely classical point of view, we would be tempted to say that the appearance of macroscopic differentiations or distinctions is not random or unpredictable: it is caused by certain microscopic differences between components of the system.

For example, in the Bénard phenomenon, what causes the water molecules in one region A to move upwards, and those in another region B to move downwards? Before the water was heated, the water molecules both in A and B moved more or less homogeneously in all directions. The heating now creates a "*non-equilibrium boundary condition*": the conditions (temperature) at the boundaries (bottom and top surfaces) of the system are such that it is impossible for the system to maintain a thermodynamic equilibrium, i.e. a macrostate with maximal but constant entropy. This means that the system becomes *unstable*: a very small fluctuation can be greatly enlarged and drive the system into a new regime. For example, if the concentration of molecules moving upward in A was a little larger than the concentration in B, then this effect will be enhanced until all molecules in A move upwards, and all molecules in B move downwards. Classically, we have a causal process:

(c, c') = (**upward concentration larger in A, upward concentration larger in B**) →
(a, a') = (**all molecules upward in A, all molecules upward in B**)

However, the distinction (c, c') is that small that it cannot be observed, or predicted from previous observations; it belongs to the (unknown) microstate of

the thermodynamic system. The distinction (a, a'), on the other hand, is clearly visible, and belongs to the (observed) macrostate of the system.

The characteristic of such an instability is that initial conditions which are very close to each other (i.e. microstates which are macroscopically indistinguishable) can give rise to final conditions which are far apart (i.e. microstates which belong to different macrostates, hence are macroscopically distinguishable). This phenomenon is sometimes called "*bifurcation*": at some point the (deterministic) trajectory of the macrostate bifurcates, so that the system must choose between distinct macroscopic trajectories (e.g. one characterized by a, another by a').

This is the point where the deterministic macrorepresentation becomes stochastic: different possibilities for further evolution are open. The path the system will choose depends upon something which is unobservable. Hence, we can only try to attach probabilities to the different possibilities, we cannot predict with certainty.

This is again analogous to the quantum observation process. The non-equilibrium boundary conditions play the same role as the coupling to the macroscopic observation set up: both enhance or magnify certain microscopic distinctions. On the other hand, they blur or erase other microscopic distinctions.

For example, certain small differences in movement of molecules within the cell A are wiped out, all molecules in A undergo the same movement. This is sometimes described as "the appearance of *long-range correlations*" between parts of a previously chaotic system.

From this viewpoint, the irreversible process described in quantum mechanics by the collapse of the wave function is seen as a self-organization of the coupled system (micro-phenomenon + macro-observation apparatus), leading to the appearance of macroscopic distinctions. The indeterminism of the observation result corresponds then to the bifurcation phenomenon, where the self-organizing system must choose between distinct macroscopic states. The incompatibility of observational couplings corresponds to the fact that different boundary conditions will generally produce different dissipative structures. Finally, the "non-local correlations" between observations are similar to the "long-range correlations" in dissipative structures.

9.4 Autonomy and adaptation

In dissipative structures the self-organization depends upon the non-equilibrium boundary conditions, i.e. on the coupling of the system with the environment. If these conditions change, the structure of the system will generally be destroyed. In this sense dissipative structures are not very stable. The reason is that the "forces" which create and maintain the internal distinctions are external themselves. If the heating of the water layer is interrupted, the Bénard pattern of cells will disappear.

In the example of quantum observation, the collapse of the wave function, i.e. the creation of a macroscopic distinction, is "forced" upon the quantum system by the external observation apparatus. If the system is dissociated from the apparatus, its state trajectory will in general leave the subspace of eigenstates corresponding to the observed distinction. (Remark that if the observation would be carried out continuously, the state of the system would not change at all. The wave function would be projected continuously onto the same eigenstate. This is called the "Zeno paradox" (Misra and Sudarshan, 1977)).

For a dissipative structure to become stable, you need an internal "*control*" which maintains the internal distinctions by eliminating or counteracting all fluctuations and perturbations which might destroy it.

An example of such a stable self-organizing system is a living cell. The internal control is here effectuated by the DNA, which selects chains of amino acids to form the structure proteins and enzymes needed for the proper functioning of the all. These proteins and enzymes are sooner or later destroyed by chemical reactions. However, there is always a recuperation of their amino acids, together with an input of new amino acids by feeding and digestion, so that the DNA can reconstruct the original polymers. Thus the organization of the cell, i.e. the interplay of distinct classes of proteins and enzymes, is maintained.

Remark that the DNA itself is not invariant: it is itself produced and reproduced by the proteins it produces. In this sense the process is circular. This phenomenon may be called "*autopoiesis*", i.e. self-production (Maturana and Varela, 1980; Varela, 1979). An autopoietic system is a system whose internal processes produce the components needed for the internal processes to occur. Thus the system maintains an invariant organization, even though the components, or substances, which carry this organization change continuously.

This is analogous to the functioning of dissipative structures: for example, in the Bénard phenomenon the distinction between the cells is invariant, even though the water molecules moving inside the cells are continuously replaced. This should remind us that a distinction belongs to the category of information, not to the category of matter, i.e. it is independent of its physical substrate (cf. section 1.4), which may change, whereas the distinction remains invariant.

What distinguishes an autopoietic system from a general dissipative structure, is that one of the distinctions produced by the system is its own *boundary*, i.e. that what separates the system from its environment. This allows the system to become more or less *autonomous*, i.e. independent of the details of the outside situation. The system creates in a certain sense its own "boundary conditions", in such a way that its internal organization (dissipative structure) becomes independent of the environment. For a living cell, this topological boundary corresponds to the membrane, which encloses the protoplasm where the internal processes occur.

In order to maintain stable "boundary conditions" the internal processes must counteract or compensate all external perturbations. This type of dynamic stability is sometimes called "homeostasis". A very simple example is provided

by the thermostat, discussed in section 2.3. If the outside conditions change (e.g. the temperature goes down), an internal process will be triggered that corrects the deviation (e.g. the heating element is activated), until the ideal boundary condition is restored (e.g. the temperature reaches 24°C).

In other words, for a self-organizing system to be autonomous, it must be able to adapt to all external changes. This allows it to maintain a stable boundary, i.e. a distinction between inside (system) and outside (environment).

We might be tempted to see adaptation, which is characterized by the continuity of the system-environment boundary, as a simple causal process. However, the conservation of the inside-outside distinction in autopoietic systems is the result of a number of internal processes which do not conserve distinctions. This becomes immediately clear if we remember that "dissipative structures" are characterized by dissipation, hence production of entropy.

This may be contrasted with other systems for which the boundary is invariant. Let us once again consider our billiard-ball example. The movement of a billiard-ball is a causal process: it conserves the distinction between inside and outside of the ball. This allows us to represent the ball as a classical "object". However, this is only due to the fact that the molecules forming the ball are rigidly connected, so that it is very difficult to separate them. There are no internal processes of adaptation or self-production.

We thus find two classes of "objects", i.e. systems with a conserved boundary: simple, mechanical systems (i.e. particles or rigid bodies) like the ball, and autonomous or adaptive systems, like an organism.

We have already stressed the interdependence of adaptation and cognition (section 2.2): in order to adapt, a system must be able to perceive and to understand the signals from the environment, and to decide between alternative ways of reacting. Hence, the information and/or entropy exchanges between autopoietic system and environment can be interpreted as rudimentary cognitive processes.

This corresponds to Maturana's and Varela's (1980) definition of the cognitive domain of an autopoietic system as the domain of all interactions in which the system can enter without loss of its identity (i.e. its self-environment distinction). However, Maturana and Varela do not wish to describe these interactions as "information exchanges". The way we have defined information transferring processes, however, is general enough to incorporate Maturana's and Varela's "perturbations and compensations".

9.5 Perception and problem-solving as irreversible processes

The first step in a general cognitive process (cf. section 2.6) is perception. Perception is basically a distinction of the features of the environment which are meaningful, that is to say which are *relevant* with respect to the system's global strategy for survival (i.e. maintenance of its self-world boundary).

This means that perception is primarily a *filtering* of the incoming information, so that only the important distinctions are retained and magnified. If all information, i.e. all physical distinctions or differences, were conserved, the internal representation of the outside situation would become so complex that the system would be unable to plan or to prepare appropriate reactions (cf. section 6.2). Therefore, it would be unable to adapt, i.e. to maintain its autonomy.

On the other hand, a mechanical system, such as a billiard ball, reacts causally to all external influences. This means that all distinctions characterizing the perturbation are conserved in the reaction of the system (cf. also section 7.3). Yet the system "survives" most interactions. So, why does a mechanical system have no difficulties with information complexity?
The reason is that a mechanical system is not self-organizing, there are no autonomous processes occurring inside the system. This means that all its reactions are immediate, and completely determined by the input. It cannot learn from past events, or anticipate future events. This implies that the domain of interaction in which the system can enter without loss of identity, is a priori fixed: if the temperature becomes too high, the ball melts.

An autonomous system, on the other hand, will generally anticipate the effect of a raise of the temperature (e.g. produced by an approaching fire) and take counteraction (e.g. running away, or cooling itself). In order to do that, it needs a filtered representation of the information provided by the environment. It is just this filtering (i.e. elimination of irrelevant distinctions), which guarantees that the system remains autonomous, i.e. independent of the causal influences of the environment.

This filtering of distinctions can occur in different, incompatible ways. A traditional example is found in what psychologists call "*gestalt switches*". This phenomenon is exemplified by certain drawings which, when looked at in one way, are recognized as representing a certain object, e.g. a rabbit, or a cube seen from below. However, if the same figure is looked at again, it is suddenly recognized as representing a quite different situation, e.g. a duck, or a cube seen from above.

The two different ways of perceiving the same pattern are called "gestalts". They are mutually exclusive: it is impossible that the same figure is interpreted simultaneously as a rabbit and as a duck. Yet both interpretations are equally plausible, and if one looks long enough at the drawing he will experience a discontinuous shift from one way of perceiving the pattern to another way.

This should remind you of the incompatibility of certain propositions or observations in quantum mechanics. Quantum observation and everyday perception are indeed two instances of the same basic type of information processing, where certain external distinctions are magnified, while others are wiped out or filtered.

One way of perceiving may magnify a distinction (p, p') and eliminate a distinction (m, m'). Another perception, however, may retain (m, m'), while filtering out (p, p'). Therefore, we may find that the same phenomenon is

sometimes categorized as m (e.g. large momentum, or a rabbit), and at other times recognized as p (e.g. a specific position, or a duck).

Classically, we would conclude that the phenomenon should be represented by the conjunction of p and m. This would correspond to the interpretation that a particle has both definite position and definite momentum, or that the drawing represents both rabbit and duck. However, we know that in practice it is impossible to observe or perceive both features at the same time.

This can be explained by the basic principle that a representation can never be as complex as the thing it represents. If all potentially perceivable distinctions would be simultaneously represented internally, then the representation would be too complex. Therefore, the process of perception is necessarily irreversible, i.e. it does not conserve the complete information.

The same reasoning can be made for the other basic type of cognitive process: problem-solving. If the internal representation would be perfect, i.e. isomorphic to the external world, then the search for problem solutions could be deterministic. The different steps in the problem-solving process would simply reflect the different changes in the outside world. In that case, however, the internal feedforward process could not anticipate, i.e. go faster than the external evolution of the world (cf. section 6.2).

An additional argument could be that, as we have shown in sections 8.4 and 8.5, the structure of time itself is determined by the interconnections between information transferring processes. Since the causal connections inside the representation would be completely isomorphic to the causal connections in the outside world, the passage of time inside the representation would be in phase with the passage of time outside. Therefore, the inside processes could never go faster than the outside processes they are supposed to represent.

Since the internal representation, by definition, cannot be a perfect image of the world, we cannot expect to find a deterministic procedure for designing perfect predictions of outside phenomena. Every internal feedforward process will be characterized by "trial-and-error", and hence will depend on (internal and external) feedback in order to correct its course.

This is clearly exemplified by the heuristic problem-solving paradigm used in AI (cf. section 1.5): for every state of the problem representation there is a choice between different operators to be applied. The only guide is an evaluation function, but the maximum of the evaluation function is not necessarily the best alternative for solving the problem. Typically one tries to follow a path initially characterized by high values of the evaluation function. After a few steps, however, it appears that the path leads to a dead end: all remaining state transitions from the last state correspond to a spectacular decrease of the evaluation function. In this situation the most reasonable course of action is "backtracking", i.e. going back to a previously reached state, and trying an alternative operator, whose initial evaluation is less than the value of the operator leading to the dead end.

Clearly, the search through the problem space is not a classical causal process. Different initial states can lead to the same final state, whereas the same

initial state can lead to different final states. In other words, the problem-solving process too does not conserve distinctions.

9.6 Learning and discovery as changes of representation

Until now, we have assumed that the representation structure which guides the perception and problem-solving processes, is invariant. The state, i.e. the activation pattern, of the representation undergoes irreversible transformations, but the representation structure is fixed. The prototypical example of such a fixed representation structure is the DNA pattern, which guides the processing of amino acids and proteins.

In higher order autonomous systems, however, there is also a non-fixed component in the knowledge needed for adaptation. This flexible knowledge is generally embodied in the central nervous system, and in the immune system (Varela, 1979). The evolution of this knowledge under the influence of environmental input may be called "*learning*". When the cognitive change takes place independently of new external input, we will speak about creativity or "*discovery*".

For example, I can "learn" to drive a car by listening to someone else's advice, and by experiencing the feedback provided by a real car, when I try to make it move. On the other hand, by reflecting internally, I can "discover" a new law of physics. In practice, it is difficult to separate learning and discovery: every cognitive change is partly caused by external stimuli, partly by internal processing.

Let us analyze some of the basic psychological paradigms used for modelling learning processes.

The simplest types of learning are "*habituation*" and "*sensitization*" (Wickelgren, 1977). If a certain stimulus occurs again and again, without announcing anything important with respect to my global survival strategy, I shall learn to ignore it. For example, if I live in a house near to a railway, I shall get used to the sound of the trains, i.e. I shall not longer notice when a train passes. I shall get habituated to the stimulus. This means that I shall not longer distinguish (consciously) between the sound of a passing train, and the absence of this sound. Therefore, habituation is merely the disappearance of an internal distinction, used to interpret stimuli.

Conversely, sensitization corresponds to the appearance of a new stable distinction. This occurs when some difference between two types of stimuli, which I did not notice before, turns out to be important for my survival strategy. For example, if I discover that there are two types of apparently similar mushrooms, of which one type is edible, whereas the other type is poisonous, I will learn readily to distinguish between both types. Similarly, a new-born child will learn rapidly to distinguish the face of its mother who feeds him from the faces of other persons, who do not.

Related to sensitization, is the process which is called "*concept learning through examples and counterexamples*". The best way for a child to learn the

meaning of the concept "dog", is by being presented to different types of animals, some of which are designated as dogs, while others are designated as "non-dogs", i.e. cows, sheep, cats,… By being confronted with all these examples, the child will learn to assimilate all four-legged, hairy creature which bark, into one category, and to distinguish it from other, similar categories.

A more complex type of learning occurs when a relation is to be established between different categories of stimuli. The simplest paradigm for such learning process is called "*association through conditioning*". If a dog is presented with food each time he has heard the sound of a bell, he will learn to associate the sound with food. This means that he will begin to salivate when he hears the sound, even when no food follows. This can be explained by assuming that a kind of "connection" has been established in the dog's internal representation between the phenomenon of hearing a bell, and the phenomenon of receiving food.

We have seen two basic types of connections between distinctions: either the connection is logical ("bell" implies "food", "no food" implies "no bell"), or it is causal ("bell" produces "food", "no bell" produces "no food"). If the connection obeys both conditions, the relation between the two distinctions is one of equivalence, or perfect correlation: no phenomenon occurs without the other one.

Formally this can be represented as follows:

logical relation: (b, f') → (f, b')

causal relation: (b, b') → (f, f')

equivalence relation: (b, b') ↔ (f, f')

Furthermore, there are a continuous number of probabilistic or stochastic connections characterized by different conditional probabilities of the type:

P (f | b), P (f' | b), P (b' | f), …

We may assume that a general process of association formation can be modelled by the continuous evolution of such a set of probability functions until an "attractor state" is reached corresponding to one of the basic types of connections (i.e. a state for which most of the conditional probabilities would be zero).

It should be noted here that in practice association and sensitization go together. In other words, it is not so that you should first have two distinctions, before you can form an association between them. For example, the formation of the distinction (mother, other person) for a baby, will proceed in parallel with the formation of the association:

(**mother, other person**) → (**feeding, no feeding**)

Another example of this interdependence can be found in the domain of scientific discovery. The concept of "baryon number" was introduced in elementary particle physics to explain certain observed regularities in elementary particle interactions. Certain reactions between particles which could be expected to occur, were empirically never found. Therefore, one postulated a conservation principle, which could constrain possible reactions so that only the observed reactions would be allowed. In order to do this one needed to introduce some property which could be conserved. This property was called the "baryon number".

One could now distinguish different collections of particles according to their total baryon number. If the baryon number of the two collections (channels) was different, the transition from one collection to the other one could be excluded. This corresponds to a causal relation, where the distinction between equivalence classes characterized by different baryon numbers is conserved.

By attributing on a purely conventional basis a baryon
number to a certain type of particles, the baryon numbers of the other particles could now be computed from the observed reactions. In this sense, the actual value of the baryon number for a particle is purely a matter of convention. The only thing which is physically meaningful is the distinction between particles with different baryon numbers, because this distinction appears to be conserved in actual interactions. The discovery of this distinction follows from the assumed existence of a causal relation.

CHAPTER 10: A summary and discussion of the previous results

10.1 Different formulations of the research problem

We began this exploration by observing that the present (r)evolution of science and society demands a framework for representing change, which is more general than the framework provided by classical science, and we set out to investigate how such a framework could be constructed. This first formulation of a research problem was very ambitious, but also quite vague, so that it could be doubted whether anything like a precise, scientific analysis could be used to approach it.

Therefore, the first step of the investigation had to be a refinement of the concepts involved in the problem formulation. The fundamental concept appeared to be the concept of "*representation*". This was studied in different contexts: physics, mathematics, systems theory, philosophy and computer science. It appeared possible to synthesize the different meanings of the "representation" concept by means of the concept of "*adaptation*". An "*adaptive representation*" was then defined as the abstract structure through which an internal, cognitive system (self) could efficiently anticipate, and hence adapt to, the external physical changes taking place in the environment.

In this sense, a representation forms a kind of abstract interface, regulating the interactions between system and environment, in such a way that the system is able to maintain a stable identity. This process of regulation or steering can be conceptualized as an internal processing of information coming from the environment. The actual information being processed corresponds to the "state" of the representation; the more stable knowledge which controls this processing corresponds to the "structure" of the representation.

Although this analysis allowed us to understand the aim or function of a representation, the actual mechanism through which this function is fulfilled remained vague. Since the cognitive mechanisms used by most adaptive systems are very difficult to observe, we looked for a domain where these mechanisms are more explicit. Science was recognized as an attempt to construct such explicit representations. The scientific method for making the representation structure and its feedforward mechanism more explicit was called "formalization"; the method for making the information exchange between representation and environment, and hence the feedback mechanism, more explicit was called "operationalization".

However, after looking at the often discontinuous evolution of scientific representations and at the difficulties of communication across the boundaries of representations or disciplines, we had to conclude that there is no universal representation to which all existing theories can be reduced. Therefore, (implicit or explicit) representations appear to have only a local and temporary value; they cannot be integrated in one absolute, global representation.

But then, if representations are relative and changeable, it becomes interesting to study just how different representations can be related, and how they can evolve. A theory of such relationships and dynamics, however, would just be another adaptive representation. But this time the objects to be represented would not be outside, physical phenomena, but representations themselves. Therefore, the theory we wish to construct would be an adaptive representation of adaptive representations, or, in other words, an *adaptive metarepresentation*.

At first sight, this new problem formulation is quite different from the original formulation, where we set out to find a representation of the most general types of change, not of the most general types of representation. The key to elucidate this apparent confusion could be found in an analysis of what we called "the classical representation frame".

10.2 Lessons learned by analyzing classical representations

If we wish to generalize the representation frame provided by classical science, we should first try to understand how this frame is structured, and which kinds of phenomena are easy, respectively difficult to represent within such a structure. Such a reconstruction of both explicit and implicit representation mechanisms used in classical science would provide us not only with a detailed view of the way a (classical) representation functions, but also with some useful ideas about how the problem domain of such a representation could be extended.

This reconstruction was carried out in a hierarchical manner: starting from the most primitive elements, and gradually adding levels of organization until a complete, coherent and self-sufficient whole came out. The basic elements, which represent distinct external phenomena, were called "objects". Primitive expressions of the representation can be constructed by coupling these "objects" to "predicates", representing properties or relations between objects.

By combining these expressions by means of conjunction and negation, we get a set of compound expressions or propositions, whose logical structure is that of a Boolean algebra. The atoms of this algebra correspond to classical states. Every proposition of the algebra can be decomposed as a union of mutually orthogonal atoms. Therefore, we have the basic property that for every proposition a, and every atom s, either s implies a or s implies the negation of a. This allows us to reduce the algebra of propositions to its subset of atoms, which forms the classical "state space".

Change can now be represented by a trajectory in the state space of the represented system. A trajectory is a continuous, parameterized path. Therefore, we must introduce a parameter, representing an absolute, linear time ordering, and a topology on state space, which defines continuity.

In order to represent the way such a change of state can be brought about, we must introduce operators, which can be composed. The linear structure of

time and the equivalence class structure of state space entail a group structure for the operator algebra.

In order to decide which operator should be applied to a particular state, at a particular time, we need dynamical constraints. These can be categorized as conservation principles, which demand that certain properties of the state be conserved during dynamical evolution, and variation principles, which demand that possible state transitions be evaluated according to a certain criterion, so that the "optimal" one can be selected.

After this structural analysis of the classical frame, we had to analyze the world view entailed by this particular structure. The formal properties of the representation could now be related to general, philosophical concepts and principles on which the classical ontology and epistemology are based. The basic concepts are: atoms, matter, energy, space and time. Some principles governing the interrelation of those concepts are: reductionism, determinism, causality, reversibility, absoluteness and rationality.

Starting from this detailed analysis, we could now try to contrast classical and non-classical representations. In order to do this efficiently, we needed a simple criterion for determining the "classicality" of a representation. Such a criterion was found by going back to the most fundamental mechanism used in a representation: the process of distinction.

We had already noticed that information processing in a representation consists basically of categorizing and selecting certain combinations of categories. The Boolean logic and causal dynamics of the classical frame provided a very simple model of such a mechanism. The model is characterized by the fact that the distinction between a category and its complement, or a proposition and its negation, is absolutely invariant. This applied as well to logical as to dynamical operators or relations. We could then define a "non-classical" representation as a representation whose distinctions are not invariant.

10.3 The correspondence between classical metarepresentations and non-classical object representations

From this point of view, we can now show that the two formulations of the research problem we have used, are basically equivalent. If first, we look for a representation frame able to represent more general changes than the classical frame, then this can be interpreted as a search for a representation frame able to represent processes where not only states, but also distinctions change. Second, the search for an adaptive metarepresentation can be interpreted as a search for a representation of the way the basic components of a representation, i.e. distinctions, evolve.

The confusion between "representation" in the first case, and "metarepresentation" in the second case, can be eliminated by demanding that the representation we are looking for be explicit. Explicit or "scientific" representations must have a structure which is sufficiently invariant to be

transmitted faithfully between communicating members of a scientific community.

This means that two scientists using the same concept, e.g. "mass", should make the same distinctions when trying to categorize phenomena, according to this concepts. For example, they should use the same formal definition of "mass", and the same measuring units and instruments in order to distinguish between objects with a different mass. With such a convention the purely subjective meaning or experience of the phenomena can be eliminated from the intersubjective representation.

For example, the same color can be experienced in a quite different way by two individuals (e.g. one of the persons might be color-blind). However, if both individuals agree to make the same distinction between one class of colors, which is called "green", and another class of colors, which is called "blue", there is no obstacle to communicate efficiently about this phenomena. The "meaning" transmitted through such communication is purely determined by the explicit, intersubjective scheme of distinctions. However, this meaning is generally different from the subjective meaning which is determined by a variable scheme of implicit distinctions.

We may consider that a really intersubjective representation must be based on invariant distinctions. In other words, it must have a classical structure. (Of course, this is only an ideal case. In practice, scientific representations are only partially explicit or intersubjective. Hence, their structure will only approximate the classical ideal.)

But how could we then construct an intersubjective representation of processes which do not conserve distinctions? The answer is simple: by introducing "*metadistinctions*". Such a metadistinction would be itself invariant, but distinguish between other distinctions which could be variable. A scheme of invariant metadistinctions would form a classical metarepresentation. This would represent, depending upon the interpretation, either the changes in classical (i.e. distinction-conserving) object representations, or general non distinction-conserving processes. In order to see what this means in practice, it suffices to look at some examples.

Quantum mechanics is an explicit representation characterized by a non-classical logic, representing the non-conservation of distinctions during the observation process. The only way to express this non-classical logic is by using the language of mathematics, i.e. set theory, which forms the base of the Hilbert space geometry and its algebra. But this language is based purely on classical, Boolean logic. This fact has often led to confusion amongst researchers in quantum logic: on the one hand, to formalize quantum theory you use a classical logic; on the other hand, if you analyze the epistemological implications of this formalism, you find a non-classical logic; which one of those in now the true logic of quantum mechanics?

In order to evade this paradox it suffices to distinguish two levels of description: 1) the classical logic of set theory functions on the metarepresentational level, it represents the structure of the Hilbert space

representation; 2) the non-classical logic functions on the representational level: it represents the external process of observation. Therefore, we find a classical metarepresentation coupled to a non-classical object representation.

A similar construction can be found in relativity theory. On the object level, relativity theory is not a classical representation: there is no invariant distinction between time (precedence) and space (simultaneity). However, this can be reformulated so that we get a classical representation on the metalevel.

An example of such a classical representation of general relativity is the theory called "geometrodynamics" (Misner, Thorne and Wheeler, 1974; Wheeler, 1968). Here we have a particular, space-like hypersurface in 4-dimensional space-time, and consider it as a particular state of the geometry of 3-dimensional space. The complete structure of 4-dimensional space-time, which is determined by the Einstein equations, can now be seen as a dynamical sequence of such 3-dimensional space geometries.

Hence, we have reduced the relativistic spatio-temporal representation to a classical trajectory in state space. However, this state space does not represent the possible properties of a concrete physical phenomenon. A state of this space (which is sometimes called "superspace") represents a complete space, with a proper geometrical structure. In other words, the states of the geometro-dynamical representation are themselves complete representations of a 3-dimensional state space. Therefore, geometrodynamics is a classical metarepresentation, representing a non-classical object representation.

The examples we have seen of classical metarepresentations of quantum or relativity theory are, however, quite incomplete. The metarepresentation of quantum mechanics does not say anything about dynamics, whereas the metarepresentation of general relativity seems to ignore all logical structures. It is no wonder then that there seems to be no basis to unify quantum mechanics and relativity theory, or even that both theories seem to be inconsistent.

In order to integrate different non-classical approaches we need a classical metarepresentation which is at once simple and general. Instead of representing merely some particular substructures of its object representation (e.g. the logic of observations, or the geometry of 3-dimensional space), it should represent the fundamental building blocks with which all adaptive representations are constructed: distinctions, and all the ways they can be interrelated, or evolve. To prepare such a metarepresentation is really the objective of the present study.

From our analysis of the classical frame we learned some basic properties of the static or logical relations between distinctions. By analyzing non-classical representations we should be able to learn something about the dynamics of distinctions. Let us review the principal results we have obtained.

10.4 Lessons learned by analyzing non-classical representations

A first striking feature of quantum mechanics is what we called the "complementarity of representations": it appears impossible to give a complete

classical description of a quantum phenomenon; we can only find mutually incompatible, partial representations. Every partial representation consists of a set of (compatible) observables, which give determined results for a particular set of states (the eigenstates of the commuting set of observables) of the quantum phenomenon. However, a basic feature of the quantum formalism is the "superposition principle", which posits that for every two (orthogonal) eigenstates s_1 and s_2 of a partial representation, there exists a "superposition state":

$$s_3 = s_1 + s_2$$

which is not an eigenstate of this representation. It is always possible to find an observable or representation for which s_3 is an eigenstate, but this representation will be incompatible with the previous one.

The simplest way to express this complementarity principle, and its formal equivalent, the superposition principle, appeared to be the introduction of a non-trivial orthogonality relation in the quantum state space. The superposition state s_3 could then be identified with a state which is not orthogonal to s_1, nor to s_2, but orthogonal to all states orthogonal to s_1 and s_2. The propositions of the quantum representation could be identified with the orthogonal closure of subsets of the state space.

The non-orthogonality of states s_2 and s_3 implies a non-zero transition probability $P(s_2 \mid s_3)$. Since the conjunction of s_2 and s_3 can never be true, this is in contradiction with the Bayes axiom of classical probability theory. This apparent paradox could be resolved by postulating that the quantum states s_2 and s_3 correspond to non-disjoint subsets of unobservable "infrastates". The transition probability P would then be proportional to the number of infrastates in the intersection of those subsets.

We still had to find an interpretation of these infrastates which would explain why they are unobservable. A quantum state, such as s_2, was interpreted as the maximal information, or amount of distinctions, which could be simultaneously magnified by an observation apparatus. However, this information can never be complete: there is always a loss of distinctions during the observation process.

The reason for this is that the microstate of the apparatus, which determines the state of the quantum system after the measurement, is itself indeterminate. This implies that those distinctions which were not magnified by the apparatus, are perturbed in an uncontrollable way by the observation, and hence can no longer be magnified by a subsequent observation. The indeterminacy of the apparatus microstate can be traced back to the principle of the impossibility of self-determination: the macroscopically observable part of the apparatus can never represent as much information as the apparatus contains as a whole, since a part can never be as complex as the whole.

The conclusion is that it is impossible to distinguish two non-orthogonal quantum states by a single experiment, they can only be distinguished statistically.

On the other hand, if a quantum system in a state s_1 is subjected to the observation of a proposition a, which is incompatible with, or non-orthogonal to, s1 (i.e., it is not so that $s_1 < a$, or $s_1 < a'$), then a new distinction is created: one part of the ensemble of systems in state s1 will be projected onto an eigenspace corresponding to a, the other part will be projected onto the eigenspace corresponding to a'. The difference between the two subensembles is due to a different microscopic relation between the system and the measuring instrument. In this sense, the distinction between a and a' is created by an unobservable difference between observation set-ups; it is not inherent in the quantum system.

If this unobservable difference is interpreted as a "hidden variable", then this variable is not localized in the quantum system. This explains why quantum coincidence experiments do not obey the Bell locality condition which demands that the results of simultaneous observations on spatially separated parts of a system would be determined locally, hence would be independent of each other.

In order to elucidate this locality problem, we had to analyze the structure of space-time and the way it is dependent upon causality. The relativity principle for reference frames, together with the existence of an invariant limit speed for causal propagation, entail the relativity of simultaneity. This signifies that it becomes impossible to determine the absolute time at which spatially separated events occur. Indeed, to measure this time we need to synchronize clocks located in different parts of space, and this synchronization requires the exchange of signals. But since these signals require a finite time interval to travel from one clock to the other one, and since this interval cannot as yet be measured since the clocks are not yet synchronized, we must arbitrarily choose the value of this interval in order to make a synchronization. Different choices entail different synchronizations, and hence different simultaneity and precedence relations.

However, the existence of an invariant maximum speed allows us to determine invariant precedence relations: horismotic precedence and chronological precedence. The first relation is established between events which can be connected by a signal travelling with the speed of light, the second between events connected by a slower-than-light signal. The union of both relations is called causal precedence. It is a temporal order relation, which is however not linear. This means that in general, two events cannot be temporally ordered. On the other hand, the incompleteness of the relation allows us to define a non-trivial topology on space-time. This topology guarantees the continuity of causal processes without requiring any additional assumptions.

In order to understand this mathematical property of causal precedence on a more fundamental level, we had to further analyze the causality concept. Locally causal connections were defined to be relations between events which conserve distinctions. In order to determine whether such a connection could be used to effectively transmit information, we had to extend this concept to globally causal connections, i.e. we had to consider sequences or paths of locally connected events.

If such a path is part of a cycle, then the connection cannot be used to transfer information. Indeed, either the indication coming back to an event after following a cyclic path is consistent with the original indication, or it is not. In the first case, the indication of the event implies itself, hence nothing has changed. We have a tautology: a = a. In the second case, the indication determining the event contradicts itself: a = a'. Therefore, the distinction (a, a') is deleted, and the causal connection annihilates itself. This corresponds to a causal paradox.

If the causal path is acyclic, it can be interpreted as an information transfer. There remain two possible cases: either the path is unique, and then it can be interpreted as defining a horismotic relation, or it is not, and then it can be interpreted as defining chronological connection. In this way the causal and hence topological structure of space-time, is reconstructed by demanding that potentially causal connections be able to transmit information.

This construction could then be used to elucidate some non-locality paradoxes in quantum mechanics. These paradoxes are characterized by a correlation between spatially separated but coincident observations. Since the correlation is a symmetric, hence cyclic, distinction conserving connection, it cannot be used to transfer information. Therefore, there is no contradiction with the locality principle derived from information transfer.

An analysis of the Aerts thought experiment with the two-vessel system has shown that it is not necessary to attribute the correlated observation results to a "common cause", inherent in the system. In this experiment, the observed indications are "caused" by the incompletely controlled interaction between system and observation instrument, and their correlation is due to a global conservation principle.

From this point of view, it is interesting to re-examine the relation between quantum mechanics and relativity theory. Both are non-classical representations characterized by the (implicit) acknowledgment that distinctions are not necessarily conserved. However, their focus of attention is different. In relativity theory the focus is on those processes which do conserve distinctions. These processes are called "causal". By distinguishing them from processes which are not causal, one succeeds in deriving certain formal properties, which can be used to structure space-time.

In quantum mechanics, on the other hand, the focus is on processes which do not conserve distinctions. They are described as "the collapse of the wave function". The distinction between such non-distinction conserving processes and classical observation processes where all distinctions are conserved, entails a distinction between orthogonal and non-orthogonal pairs of states. This "metadistinction" can be used to structure the quantum state space. This gives rise to the superposition principle and to the Hilbert space structure.

In this way we find two complementary descriptions: on the side of the distinction conserving processes we find relativity theory, characterized by the causal space-time structure which is determined by an incomplete precedence

relation; on the side of the non-distinction conserving processes we find quantum mechanics, characterized by the Hilbert space structure, which is determined by an incomplete orthogonality relation.

From this viewpoint, quantum mechanics and relativity theory are not inconsistent or incompatible: they only describe the two opposite sides of the same phenomenon. What distinguishes them both from the classical representation frame, is that they acknowledge that there are two sides. In the classical world view, only one side is assumed; the possibility of non-distinction conserving processes is simply ignored.

Moreover, if the analysis is carried through far enough, the origins of the non-classicality in relativity theory and in quantum mechanics seem to be quite similar. Both the superposition principle and the causality principle can be viewed as instantiations of a general principle of the impossibility of circular information transfer. In the relativistic case, there is no information transfer from an event to itself. In the quantum case, there is no information transfer from an observation instrument to itself. A further study of this analogy seems very promising.

The main obstacle to such an examination appears to lie in the great difference between the formalisms and concepts used in relativity theory, respectively quantum mechanics. The concept of distinction and causal connection, have allowed us to make a first comparison between the two approaches. However, we have only scratched the surface of a vast problem domain. The unification of relativity theory and quantum mechanics demands a completely new formalism. A possible strategy to construct such a formalism will be discussed in chapter 11.

In a certain sense, the problem domains of quantum mechanics and relativity theory are still very simple. They both try to represent phenomena without internal structures: elementary particles and events. This entails that all processes involving these phenomena are external. Therefore, it is relatively simple to classify them as distinction conserving, respectively non-distinction conserving.

This is no longer true when complex systems are considered. We have characterized the evolution of such systems by the somewhat vague concept of "irreversibility". During such an evolution certain distinctions are conserved, whereas other distinctions are created or destroyed. However, it is generally quite difficult to determine which distinctions will be conserved, and which distinctions will be created or deleted. Moreover, the total amount of observable distinctions is not constant. If the system is closed, this amount will generally decrease (second law of thermodynamics).

We could further speculate that this law too can be reduced to the principle of impossibility of circular information transfer. An argument for that could be found in Brillouin's (1962) resolution of the paradox of the Maxwell demon.

According to Maxwell, this demon would be able to decrease the entropy inside a closed system, by creating new distinctions through observation. For example, he would distinguish between the two types of molecules in a mixture

of gases, and let the individual molecules of one type pass in a compartment A, those of the other type in a compartment B. After a certain time, this would create a macroscopic distinction between both compartments. This would correspond to an entropy decrease.

However, Brillouin has shown that the interaction between demon and molecule during the process of observation would create an amount of entropy which is larger than or equal to this entropy decrease. We could compare the demon to a quantum observation instrument, and consider that the gain of information during the interaction is necessarily counterbalanced by the loss of other information, because of the indeterminate perturbation of the observed system (molecule) by the apparatus. Since the demon cannot have complete self-knowledge, he cannot prevent a certain transfer of indeterminacy (i.e., entropy) from itself to the observed system. Therefore, the total entropy of the system demon + molecules cannot decrease.

Once again, we should remark that a more thorough analysis of these problems requires a more elaborated and formalized conceptual framework.

We have further discussed some examples of higher-level complex systems, characterized by self-organization, autonomy and cognition. Here the internal increase of entropy can be compensated by external input of information. However, here too we have shown that the principle of incomplete self-knowledge constrains the information intake. This leads to a filtering of information during perception. On the highest level, these irreversible information processes take the form of changes of the representation structure of cognitive systems. We had to conclude once again that the modelling of these phenomena requires a metarepresentational formalism. We will now propose a possible basis for such a formalism.

CHAPTER 11: Towards a formalization and operationalization of the theory

11.1 Introduction

In the preceding chapter we have summarized some basic principles on which an adaptive metarepresentation could be founded. We remarked that if we would wish this metarepresentation to be scientific, we should provide a scheme of invariant distinctions determining as well the structure of the metarepresentation, as its relation to the outside environment. In order to do this we must try to *formalize* as well as *operationalize* the theory.

In the present state of knowledge, however, no definitive formalism or operationalization can be proposed. We can only sketch some general principles on which a provisional formalism could be based. We can further present some procedures by which such a formalism could be applied to practical problems, so that its adequacy can be tested. By experimenting in this way, it should be possible to elaborate, refine and, if necessary, correct the formalism, until a concretely applicable theory is obtained.

This is work for the future. For the moment, it will suffice to show that the concepts which were uncovered are sufficiently well-structured and concrete to form the basis for an explicit representation.

If we look for a criterion for evaluating the adequacy of a theory, we must first define the objective of the theory. A metarepresentation was defined as a theory allowing to anticipate or plan changes of representation. Hence, its objective (i.e. the problems it is supposed to solve) should be to construct an optimal (or at least satisfactory) representation of a given problem domain. In other words, it should provide a method for transforming an incomplete, unnecessarily complex or ill-structured representation into a simple and well-structured representation. Some examples, and the first steps in the development of a computer model of such a representation transformation can be found in (Korf, 1980).

Another interpretation, and hence domain of application, of a metarepresentation is the modelling of non-distinction conserving processes. The two most interesting examples of such processes are self-organization and discovery. Some interesting work about the computer simulation of discovery (which is related to representation transformation) has been done by Lenat (1983). There are further numerous attempts to simulate self-organizing processes, and even autopoietic systems (Varela, 1979), on computer, in particular through the paradigm of "*cellular automata*" (Farmer, Toffoli and Wolfram, 1984).

The problem with existing formal or computer models of such phenomena, however, is that either they are too simple to model really interesting processes, or they are too complex to be practically managed. In order to avoid this, we

must try to construct a formalism, which can be made arbitrarily complex, by adding new distinctions, but which, on the other hand, can always be reduced to its bare essence by eliminating all redundant or irrelevant distinctions.

We will now try to build up such a formalism beginning with something similar to the hierarchical construction of the classical frame, to which we will add further non-classical levels of representation. The formal structures we used to describe, what we called "the classical frame" (e.g. Boolean algebra, group structures,…) are indeed already metarepresentations. However, they contain a lot of *redundancies*. The addition of non-classical formal structures (e.g. a non-trivial orthogonality relation, or a causal precedence relation) will lead to further redundancies, together with apparent inconsistencies.

Therefore, we must try to reconstruct all these structures from the basis of pure, *contentless distinctions,* so that all constraints would become totally transparent. No redundant or unmotivated structures should be left. For every structure (i.e. basically a set of distinctions) we should be able to decide whether it is relevant for the problem domain, or not. If it is irrelevant we should be able to delete it. If a new problem domain is considered, new distinctions will become relevant, and we should be able to add them to the representation.

No existing formalism offers these features. Every existing formalism has a set of in-built structures, determined by axioms, which must be used throughout all applications of the formalism. Furthermore, there is only a limited set of structures which can be added to the formalism in order to enhance its scope. The formalisms we have considered (Boolean algebras, partial orders,…) were already chosen such that their basic structures are very simple, whereas their domain of applicability is rather broad.

However, if we consider them all together, we get a structure which is not very simple anymore, but whose domain of application is still restricted. The complexity of the representation structure leads to confusion, to difficulties in problem-solving and to paradoxes. So, we must try to simplify the formal structures we already found, while broadening their domain of applicability.

Let us begin by studying the different ways by which (static) combinations of distinctions can be made.

11.2 Distinction algebras

The Boolean algebra, which forms the basic structure of classical logic, is defined by means of a collection of different axioms, relating the connectives (conjunction, disjunction, negation,…) to the variables, and constants (I and O) of the algebra. However, Spencer-Brown (1969) has shown that this algebra can be derived from an *algebra of distinctions*, which uses only one connective (⌐,), determined by two axioms. In order to simplify notation, I have replaced the special symbol ⌐ used by Spencer-Brown (the "mark", symbolizing the act of distinction) by square brackets: [ ]. Symbols covered by the mark are replaced in my notation by symbols enclosed in square brackets.

The axioms are:

1) [[p] p] =

2) [[pr] [qr]] = [[p] [q]] r

This can be transcribed in the notation we used for Boolean algebras (section 4.4), by observing the conventions:

[p] = NOT p

pq = p OR q hence: [[p] [q]] = p.q

The axioms then become:

1) p'.p = 0 (law of contradiction)

2) (p OR r).(q OR r) = (p.q) OR r (distributivity)

Spencer-Brown has shown how Sheffer's set of postulates for Boolean algebras, which is the least such set, can be derived from these two postulates.

The algebra of distinctions, however, is derived from an even more simple formal structure: the *calculus of distinctions*. In this calculus there are no variables denoting different distinctions. There is only one distinction considered, which could be interpreted as the distinction between "true" and "false", and which is represented by [ ]. The rules of calculation are derived from two extremely simple axioms:

i) [[]] =

ii) [] [] = []

In our notation this could be transcribed by using one variable between brackets: (a), to indicate that it is not a specific distinction:

i) (a)" = (a) (double negation)

ii) (a) OR (a) = (a) (idempotence)

The sign can be interpreted as the crossing of a boundary or distinction between two states: the marked state (represented by the inside or concave side of the sign) and the unmarked state (represented by the outside or convex side of the sign).

Axiom ii) means that if you cross a boundary two times in sequence, from the outside to the inside, and from the inside to the outside, you come back to the state you started from: nothing has changed.

Axiom i) signifies that if you cross a boundary two times in parallel, from the outside in, and from the outside in, then this is equivalent to only crossing the boundary once.

The axioms i) and ii), or 1) and 2), can be used to prove a number of theorems (Spencer-Brown, 1969). These theorems can be used to *simplify* expressions consisting of combinations of connectives ([ ]), and variables (a, b, c, ...). Since all the expressions of Boolean logic can be transcribed in such expressions of the algebra of distinctions, this formalism can be used to simplify expressions from classical logic. The ultimate simplification of an expression consists in reducing it to one of the two primitive expressions:

a) [] (i.e. "truth")

b) (i.e. "falsity")

The sequence of steps leading to this simplification can be interpreted as a *proof* of the expression, or as a proof of its negation.

In general the simplified expression will still contain variables. For example:

((p < q).(r < s).(q OR s)) < (p OR r)

can be transcribed as:

[[[[p] q] [[r] s] [qs]]]

and this can be simplified to (Spencer-Brown, 1969; p. 116):

[qs] pr

which can again be transcribed as:

(q OR s) < (p OR s)

The truth of such an expression will be contingent upon the truth or falsity of the variables q, s, p and r. However, to check the truth of the simplified expression for a given set of values for the variables is much easier than to check the truth of the original expression.

This method of simplifying or proving propositions is essentially similar to the *resolution method* used in artificial intelligence and in automated theorem proving (see e.g. Charniak and McDermott, 1985).

In Boolean algebra the method is based on the transformation of a Boolean expression to its *conjunctive (or disjunctive) normal form*. A conjunctive normal

form of an expression consists of a conjunction of factors, which themselves consist of the disjunction of a number of primitive expressions or their negations, and such that every factor contains the same set of primitive terms in positive, or negated form.

For example:

(a OR b'). (a OR b)

is a conjunctive normal form, with two factors, and two primitive expressions a and b. Every Boolean expression can be reduced to such a form, by changing all implications into disjunctions, and by applying the rules of distributivity, and the laws of de Morgan. Once you have a normal form, it is easy to delete redundant variables (i.e. distinctions).

For example, the expression a is part of both factors in the above normal form, and hence, because of distributivity, it can be taken out of the brackets. We get:

a OR (b'. b) = a (since b'. b = 0)

This means that the truth value of b is irrelevant for determining the truth value of the complete expression.

Suppose that the problem with which an adaptive system is confronted, consists of attaining a goal represented by the originally compound expression. This means that the system will have to conceive a sequence of actions which will lead from the expression representing the actual situation to the expression representing its goal. In order to do this in an efficient way, it is best to maximally simplify the expression, so that a minimal set of primitive variables is left. The system can then easily check which of the variables has already the required truth value, and which should acquire a different truth value. It can then plan a sequence of actions consisting of different subsets, which each change the truth value of one primitive variable (i.e. cross the boundary of one distinction)(this procedure is analogous to the "means-ends analysis" used in the General Problem Solver conceived by Newell and Simon, 1975). The more we succeed to simplify the expression for initial and goal state, the easier it will be to design such an action plan.

This is a concrete example of problem transformations. However, the present transformations are limited to the purely logical combinations of representation elements. We must now also try to formalize the kinematical, geometrical and dynamical properties of representations.

11.3 Kinematical constraints for distinctions

The distinction algebra of Spencer-Brown is based on the assumption that all variables have a fixed truth value. This means that they can be interpreted as different formulations of the one basic distinction between truth and falsity,

represented by ⌐¬. Their apparent variability is only provisional, because we do not know their truth value as yet.

However, to construct an adaptive representation, we need a set of different states, so that for each state there is a specific set of primitive expressions which are true, whereas the other primitive expressions are false. Therefore, for different states, different primitive expressions will be true. The space of possible states will then be determined by all possible assignments of truth values to the primitive expressions. In a free Boolean algebra, this means that every (atomic) state can be written as a conjunction of the primitive expressions which are true for this state, and the negation of the primitive expressions which are false for this state:

$s = a_1.a_2. \ldots.a_n.b_1'.b_2'. \ldots.b_m'$

Every compound expression can then be written in its disjunctive normal form, as a disjunction of terms consisting of a conjunction of primitive expressions and their negations. If the terms contain all primitive expressions and their negations, they correspond to atomic states. Hence, every expression can be written as a disjunction of atomic states.

In practice, not all primitive expressions will be independent. This means that there will be constraints, so that certain combinations of primitive expressions will be excluded. The basic relation of this type between primitive distinctions is *implication*:

$a < b$, which is equivalent to: $b' < a'$.

Remark that this can also be expressed in a notation similar to the one we used for causal relations (cf. sect. 9.6):

$(a, b') \rightarrow (b, a')$

For a causal relation we should have:

$(a, a') \rightarrow (b, b')$

Here we see that for an implication relation the two distinctions (a, a')and (b, b') are "mixed" in some way, so that the two sides of the connection expressed by → contain each states from both distinctions. This can be interpreted to signify that the two distinctions are simultaneous; there is no unambiguous order between them. The cyclic or symmetric causal relations, which were called correlations:

$(a, a') \rightarrow (b, b')$ and $(b, b') \rightarrow (a, a')$

can be interpreted as special cases of implications, namely symmetric implications or equivalencies:

$(a, b') \rightarrow (b, a')$ and $(b, a') \rightarrow (a, b')$

or: $(a < b). (b < a)$

In this sense, correlations can be viewed as relations between distinctions which are both causal relations and implications.

In a Boolean algebra implications between primitive expressions: $a < b$, will usually be expressed as identities of the form:

$a. b = a$ or $a. b' = 0$

This means that in a disjunctive normal form the conjunction a. b or a. b' can be simplified to a or to 0. If such identities hold, the Boolean algebra is not "free" (Halmos, 1963). This means that it is not isomorphic to the power set of a set whose elements correspond to the maximal conjunctions of all primitive expressions or their negations. The distinct elements of the Boolean algebra will be smaller in number than the elements of such a power set, because the constraints imply that certain combinations of primitive expressions should be identified (e.g. a.b can be identified to a).

There is another type of kinematical constraints which must be introduced. As we have seen, it often happens that certain Boolean combinations of variables are not observable, even though the primitive expressions are. The simplest way to represent this kind of limitations is by introducing a "*closure*" operation in the algebra (see section 7.2 for the properties of such an operation.). The observability of an expression can then be equated to the fact that it is closed. Non-closed expressions would correspond to combinations of primitive expressions which cannot be macroscopically distinguished.

This method can be used as well to represent a non-classical logic within a classical Boolean structure (see section 7.2 for quantum logic, and (Sikorsky, 1969; p. 199) for modal and intuitionist logics), as to represent a topological structure in a Boolean algebra (Sikorsky, 1969; p. 198).

In the latter case, a closed expression can be interpreted as a "closed" subset of the state space consisting of the atomic expressions. The set of all closed subsets defines a topology an state space. A closed subset can be interpreted as a set of states which can be unambiguously distinguished from its exterior, that is to say there is a "boundary", separating the interior from the exterior. Such "macroscopically observable" distinction could be represented by an orthogonality relation, analogous to the one used in section 7.2.

Two states would be said to be orthogonal if there is a boundary separating them. The closure operation could then be defined as taking the orthogonal complement twice:

Closure (a) = $a^{\perp\perp}$ (cf. section 7.2)

Remark that the implication relation can also be seen as an orthogonality relation between one of the expressions and the negation of the other one:

$a < b$ *iff* $a \perp b'$

From this point of view, all kinematical constraints (implications, limited observability, topological structures,…), could be represented by adding one simple type of relation to the distinction algebra of Spencer-Brown: *orthogonality*. However, the meaning of this orthogonality relation with respect to the basic structures of the distinction algebra is not very clear yet. We might speculate that states or propositions would be orthogonal if they can be distinguished directly, i.e. by means of a single observation. In order to elucidate this, however, we should first have a theory of the dynamics of distinctions.

11.4 Categories of distinction algebras

Now that we have seen some formal methods for describing the possible states of a representation constructed from pure distinctions, we should try to formalize the possible transitions between states. In general a transition or change from one description to another one can be represented by a *morphism* of Boolean (or distinction) algebras (cf. Halmos, 1974). A morphism is a mapping f from one algebra to another one, which conserves the basic structures. For a Boolean algebra, this means that we have:

$f(p.q) = f(p).\ f(q)$

$f(p') = (f(p))'$

If we would also represent kinematical constraints by using an orthogonality relation, we should add a requirement of the type:

$a \perp b$ *iff* $f(a) \perp f(b)$

If the orthogonality relation represents a topology, the requirement amounts to a form of continuity for the mapping. If it represents a quantum logic, the requirement can be interpreted as a form of unitarity for the mapping.

The restriction of general mappings to morphisms should be viewed as a requirement that a process should be minimally structure preserving in order to

be represented. A process which would conserve no distinction whatsoever, would be unable to transfer any information, and hence would be totally uncontrollable and unobservable. Therefore, it does not make sense to include it in a representation.

However, we have emphasized that we do not want to restrict our representation to classical causal processes which conserve all distinctions. Therefore, the formalism should be sufficiently general to encompass processes which are only partially distinction conserving. Instead of a transformation group, which represents classical processes, we need a more general algebraic framework. This can be found in "*category theory*" (see e.g. Arbib and Manes, 1975).

A "*category*" K is defined as a set of (formal) objects Obj(K), a set of morphisms K(A, B) going from one object A to another object B, and a composition operator ($*$) for morphisms which is associative, and for which there exists an identity morphism for each object:

Morphisms: $f \in K(A, B)$ is also written as $f: A \rightarrow B$

Composition: $\forall f \in K(A, B), \forall g \in K(B, C), \exists h \in K(A, C): h = g * f$

Identity: $\forall A \in Obj(K), \exists! id_A \in K(A, A): f * id_A = f, \forall f \in K(A, B)$

It is possible to identify each object A with its associated identity morphism (which sends the object onto itself, without changes): id_A. In that case, the category consists only of morphisms. This is sometimes called the "*arrow only*" description of a category, because morphisms are usually represented as arrows ($\rightarrow$) connecting different objects (or other arrows).

This framework can now be used to represent and classify all partially distinction conserving processes within the formalism of Boolean (or distinction) algebras. Consider a category whose objects are Boolean algebras, and whose morphisms are morphisms of Boolean algebras. These morphisms can be divided in distinct classes:

1) Morphisms which map a Boolean algebra B onto itself (f(B)=B), are called *automorphisms*. They send each distinction (a, a') of B onto another distinction (b, b') of B, and for each distinction (b, b') of B there is another distinction (a, a') which is sent upon it. An automorphism is completely distinction or structure preserving.

The set G(B) of all automorphisms of B, which is a subset of K(B, B), is a group. Indeed: there is an associative composition operation, which is internal in G(B); there is an identity morphism which sends each element of B onto itself; the inverse of an automorphism is again an automorphism. Moreover, automorphisms send atoms to atoms. Hence, the set of automorphisms corresponds to the classical transformation group of a state space (cf. section

4.7). The Boolean algebra B with respect to G(B) can be interpreted as a classical system or object.

2) Morphisms which map a Boolean algebra *into* itself (f(B) B) are called *endomorphisms*. Each distinction (a, a') of B will now be sent upon another distinction (b, b') of B, but not all distinctions (b, b') will be the image of another distinction (a, a'). We may say that the morphism sends B onto a smaller algebra f(B) with less distinctions. Hence, the morphism deletes or "forgets" distinctions.

Therefore, it represents a process where there is a *loss of information*. An example of such processes are the irreversible processes in statistical mechanics (see section 9.2) where there is a transition from a microstate description to a macrostate description, together with an increase in entropy.

The subcategory K(B, B) of all morphisms from B to B, which comprises both endomorphisms and automorphisms, can be seen as a category with only one object (B). Such categories are called *monoids*. They have the structure of a *semigroup:* there is an internal, associative composition with an identity element, but in general there are no inverse elements.

3) Morphisms from an algebra A to an algebra B which are both onto and into, are called "*isomorphisms*". There is a bijective correspondence between all distinctions (a, a') in A, and all distinctions (b, b') in B. All isomorphisms can be inverted. However, the subcategory of all morphisms from A to B, and from B to A which have an inverse, is not a group, since there is no unique identity element: there is one identity id_A in A, and another identity id_B in B.

If A and B are connected by an isomorphism, they can be interpreted as representing *perfectly correlated* objects: every distinction in A is determined by a corresponding distinction in B, and vice versa. An example of such a system could be found in the EPR set-up (see section 8.5): the conservation of spin induces an isomorphism between the distinction algebras representing spin measurements in different directions carried out on the two particles (remark that in this case the algebras are not purely Boolean, we must introduce a non-trivial orthogonality relation to represent the superposition states). Another example is a system consisting of two rigid bodies (e.g. billiard balls), which are rigidly connected to each other. The classical representations of the two bodies are perfectly correlated: there is a bijective correspondence between the two state-spaces.

4) Morphisms from A to B which have no inverse can be interpreted as signals sent from A to B. They correspond to the non-cyclic distinction conserving connections of section 8.4, which define the precedence relations. They can be further classified according to the specific distinctions they conserve:

a) a morphism which is *one-to-one into* will send different elements or distinctions upon different elements or distinctions. It is called a *monomorphism*. It conserves all distinctions of A, but if it is not an

isomorphism, it will not cover B completely (f(A) B). This means that certain distinctions of B will not be determined by A.

This can be interpreted as a "creation" of distinctions in B with respect to A. For example, if B corresponds to a dissipative structure, and A to its environment, then clearly distinctions will appear in B which are not completely determined by the boundary conditions provided by A. This is the phenomenon of "bifurcation".

b) a morphism which is *onto* will cover B completely. It is called an *epimorphism*. If it is not an isomorphism, this means that it will send distinct elements of A upon the same element of B. Therefore, certain distinctions of A will not be conserved. However, the distinctions of B will be completely determined by those of A.

This is analogous to an endomorphism: the process leading from A to B is deterministic, but accompanied by a loss of information. For example, if a complex system A emits a particle B, the particle's state will be completely determined by the initial state of the system. However, the observation of the particle's state will not provide sufficient information to determine this initial state.

c) a morphism which is neither an epimorphism nor a monomorphism corresponds to a process where there is as well a deletion of distinctions from A, as a creation of distinctions in B. This is the most general case.

5) Morphisms which send a Boolean algebra B onto the trivial Boolean algebra "2" which consists of only two elements (i.e. one distinction): I and O, are called "*2-valued morphisms*". They can be interpreted as representing a particular state of B. All propositions of B which are sent upon I are assumed to be true, all propositions sent upon O are assumed to be false. Since this morphism conserves the Boolean structures (negation, conjunction,…), the set of true propositions will not be inconsistent but will correspond to a particular maximal conjunction of propositions, denoting the (atomic) state. The transition between two states s_1 and s_2 of B, represented by two 2-valued morphisms, can then be represented by an automorphism f from B to B, such that:

$$s_2 * f = s_1$$

The possible states of different objects defined in this way can be conceived as representing potential events. The set of events can then be structured in the same way as in sections 8.3 to 8.5. The morphisms between (non-trivial) objects could be viewed as representing causal connections leading from one event to another one. For example, the morphism f above leads from event s_1 to event s_2. The sequences or "paths" of morphisms for which there is no inverse morphism, could then be interpreted as defining horismotic and chronological precedence relations. These relations would then determine a temporal order, a topology, and possibly a metric.

A minimal realization of such a relationally determined space-time structure can be found in (Heylighen, 1984). In this model there are however no explicit distinctions. This is equivalent to a model where each object is characterized by only one distinction: (presence, absence) or (existence, non-existence) of an event. The "arrows" of the "structural language" (Heylighen, 1984) can then be interpreted as morphisms which conserve this unique distinction.

If more than one distinction is considered, however, the model becomes much more complex, and the interpretation of distinctional states as events, or morphisms as processes, is much less straightforward. In order to further elucidate this problem we should gain a better understanding of the relation between morphisms and distinctions.

11.5 The relation between categorical and Boolean algebras

We have succeeded in reducing the set of formalisms used to build up a classical representation (objects, predicates, Boolean algebras, state space geometry, linear order of time, group structure of operator algebra,…) to only two formalisms: Boolean algebra and category theory. Moreover, we have shown that these particular formalisms can be generated by a very small set of postulates.

In order to represent a classical or non-classical representation within this framework, it suffices to identify an object with a Boolean algebra, a predicate characterizing this object with an indication or proposition in the algebra, the logic of propositions with the algebraic structure, the state space with the set of atoms, the operators, processes or relations applied to objects with morphisms between algebras, the events with states of objects, i.e. 2-valued morphisms, the causal structure of space-time with the horismotic and causal precedence relations determined by non-reversible morphisms,…
Of course, the details of the construction must still be worked out. We have only attempted to show that existing formalizations can be simplified and generalized through an analysis based on the concept of distinction.

A further simplification requires the elucidation of the relation between category theory and Boolean algebra. The question which should be asked is: is it possible to synthesize both formalisms, so that only one algebra is left? For the moment, we cannot give a definite solution to this problem. However, there are reasons to suppose that categorical and Boolean approaches are merely different points of view directed to the same problem.

We have already pointed out that objects in a category (in this case Boolean algebras) can themselves be regarded as morphisms, so that all of their properties can be expressed as properties of the composition of morphisms. This is typical for category theory: you do not have to specify the internal structure of the objects or morphisms in order to describe them. For example, we have defined epimorphisms as morphisms which cover all elements of their image set. However, in category theory, one uses the following equivalent definition:

Definition: an arrow f in a category K is an epimorphism *iff* the equality $g * f = h * f$ in K always implies that $g = h$

With such a categorical definition, we do not need to pay attention to the elements (i.e. the internal structure) of the objects. This is in accordance with the general philosophy underlying the present approach: knowledge, or information, is independent of its substrate or building blocks; it is only determined by abstract *relations*. (A more elaborate presentation of this "relational" philosophy of representation, where the concept of distinction is not yet used, can be found in (Heylighen, 1984)). The important point about a distinction does not reside in the specific states, objects, or elements which are distinguished, but in their relation.

We have defined a morphism as a mapping which *conserves distinctions*. However, except in the case of identity morphisms, it *does not conserve the elements* which are distinguished. But this is not important: a distinction is abstract, so is information. Therefore the morphism can be interpreted as an invariant piece of information travelling from one object to another one. The objects (i.e. Boolean algebras) themselves can be interpreted as static, or stored "pieces of information". With this interpretation a morphism could also be seen as a distinction algebra, isomorphic or equivalent to the distinction subalgebra which is its image.

Hence, we have shown that it is not only possible to interpret a Boolean algebra as a morphism, but also to interpret a morphism as a Boolean algebra. What should be done now, is to express this intuition formally by integrating the basic operations of Boolean algebra (distinction, or conjunction and negation) and of category theory (composition).

11.6 The dynamics of distinctions

A last requirement to transform the proposed formalism into an adaptive representation is the formulation of dynamical constraints, which would govern the flow of information in the model. A first type of constraint is clearly *consistency*: if an object receives information from different other objects, these informations should not contradict each other. This seems to be difficult to determine in a local way, since the different objects do not know of each other which messages they transfer.

One way to evade this problem, is by letting contradicting messages neutralize each other. This can be conceived in the same way as our discussion of causal paradoxes in section 8.4: if one object sends a message implying a, and another object sends a message implying a', then the distinction between a and a' is (momentarily) deleted from the algebra. In other words, the state of the object with respect to the distinction (a, a') becomes indeterminate. This provides us with a first sketch of a criterion for the *elimination* of distinctions.

In order to build a complete dynamics of distinction however, we would also need a criterion for the *creation* of distinctions. In a certain sense, the messages carried by different morphisms are already distinct. However, not all distinct

messages are sent upon distinct indications or propositions in a Boolean algebra. For example, the messages sent by two different objects may both actualize the proposition a. In other words, both messages are assimilated to the category a.

As an illustration, consider the new-born baby who is still unable to distinguish his mother's face from the faces of other people. All impressions of different human faces will be assimilated to the same category. However, the child will soon learn that this undifferentiated collection of messages is followed by another set of messages where there is a clear distinction: either he receives breast-feeding from the person whose face he saw, or he does not. Therefore, he will tend to "shift this distinction backwards", so that also the original set of impressions will be divided in distinct categories.

The basic requirement for introducing such a new distinction is that it be conserved by certain morphisms (representing (partially) causal relations), and so becomes connected to already existent distinctions. The general process might be described by some kind of blind-variation-and-selective-retention (BVSR, cf. Campbell, 1974). The new distinction, which is not the mere conservation of an existing distinction, can only arise in a "blind", indeterminate way. However, once it has been created it will tend to disappear in the same way, unless it is stable, i.e. conserved. This conservation or retention of a distinction depends upon the "environment" of other distinctions and morphisms, which carries out a selection. E.g. for the new-born child, the selection of new distinctions is determined by the efficiency of the new distinction in allowing to anticipate important events (feeding, or no feeding).

For the moment, these ideas about the dynamics of distinctions remain vague. Probably the best way to make them more concrete, is by operationalizing the theory, so that it becomes possible to observe actual processes happening in a system which is described by a category of Boolean algebras, and thus experimenting with different types of dynamics.

11.7 Towards an operationalization of the theory

In order to be able to apply an adaptive metarepresentation for the anticipation or steering of the behavior of a concrete system, we must find systems which use adaptive representations. The best example of such a system is the *human brain*. In principle an adaptive metarepresentation should allow us to anticipate or steer our own cognitive processes. This steering of human cognition is already carried out in a primitive form in disciplines such as pedagogy and psychotherapy. However, the human cognitive system is very complex and very difficult to observe or to analyze.

Another type of system using adaptive representations is an *organization*. Here the processes of information processing, decision making and problem solving are already more explicit, hence easier to observe and to steer. This is attempted in disciplines such as management science, or organization design (see e.g. March and Simon, 1958; Galbraith, 1977). However, the basic

components of the system are still humans, hence difficult to observe or to control.

For non-human organisms, e.g. animals, we have mainly the same problems of complexity, and limited observability.

Therefore, the best objects for testing the adequacy of an adaptive metarepresentation are artificial cognitive systems, more in particular: digital *computers*.

We have already mentioned some attempts to simulate self-organization, discovery and representation change on computers. The main difficulties experienced during these attempts, however, were not of a technical nature, but were due to the lack of a clear and general conceptual framework which could be formalized and translated as a programming language. If the formalism which was sketched in de preceding sections proves to be adequate, then it should be relatively easy to implement it on a computer. Some arguments to show that this is realistic can be found by looking at examples where structures similar to parts of the formalism were already implemented.

The programming language *PROLOG* which is used in AI, and more especially for the design of expert systems, is based on the resolution method we sketched in sect. 11.2. It can be seen as a computer implementation of the logical part of the classical representation frame. Therefore, it encompasses everything we said about objects, predicates, Boolean algebras, … However, every idea about dynamics, space or time is lacking in PROLOG. One way to evade this is by using PROLOG on a metalevel, for the determination of an object representation which may be dynamical and non-classical (cf. Heeffer, 1986).

The second aspect of the formalism, which is determined by the morphisms passing messages between objects, is similar to another AI technique: *object oriented knowledge representation*. Here the principle is to structure knowledge in modules (objects) which can communicate by exchanging data (Steels, 1984).

This general idea of communicating modules or processors is also at the base of a new development in computer hardware: "*parallel processing*", or "*connection machines*". Although technically it is possible to construct machines consisting of a network or array of parallel processors, the main difficulty is the creation of new programming formalisms, which can efficiently steer the very complex flow of information in such a machine. Two conceptual frameworks which are promising in this respect, are "semantic networks" and "cellular automata".

A *semantic network* is a frame for knowledge representation similar to a network of message passing objects (cf. Charniak and McDermott, 1985). The objects represent concepts, which are related through certain semantic relationships (inheritance of properties, causal relations, general two-place predicates,…). This structure contains no explicit dynamics: new concepts, predicates or propositions must be introduced from the outside.

A *cellular automaton* (Wolfram, 1984), on the other hand, is an abstract dynamical system, with a discrete internal space-time structure, but without explicit semantics. It consists of an array of cells, which can be in different states. There is a discrete time parameter, so that after each time unit which has passed, all cells undergo a state-transition. The dynamical law is such that the state of a cell at time t + 1 will be determined by the states of this cell and its neighboring cells at the previous time t. The cells with their (discrete) state-spaces correspond in our model to Boolean algebras. The information transfer from a cell at time t to a neighboring cell at time t + 1 corresponds to a "direct" morphism, representing a unique causal path (cf. sect. 8.4).

Therefore, both semantic networks and automata can be conceived as particular cases of the more general category of Boolean algebras. If such structures are relatively easy to implement on digital computers, then we could expect to encounter no major difficulties when trying to implement our distinction formalism.

CHAPTER 12: Formal deduction of representation structures from a metarepresentational framework

12.1. Introduction

In the previous chapters, the sequence followed was basically inductive: starting from known data (the formalisms of existing physical theories), a new, more general framework was abstracted, in such a way that the given theories (classical mechanics, quantum mechanics, relativity theory and thermo-dynamics) could be recognized as special cases of the representations within this framework, each characterized by particular features. These distinctive features were shown to be determined by the way distinctions are conserved within each representation.

In the present chapter I would like to examine in how far this sequence can be followed in the inverse direction, i.e. in how far it is possible to postulate the general framework from the beginning and then to deduce the representation structures from these postulates, with a minimum of additional assumptions.

In order to do this I shall first have to make the basic concepts and assumptions of the metarepresentational framework as explicit as possible. More in particular I shall try to indicate the axioms needed for an unambiguous, mathematical determination of the principles underlying the framework.

Of course this formalization is in the present stage of research only provisional. At present, it will only be used to eliminate possible confusions which may have arisen while reading the more informal text of the preceding chapters, and to show that the framework is sufficiently powerful to allow the derivation of some very general, yet non-trivial results, starting from minimal assumptions.

After this definition of the framework I will reconstruct the fundamental structures (i.e. the structures independent of the particular type of phenomenon being represented) of the four physical theories by incorporating their respective assumptions with respect to the conservation of distinctions.

This will be carried out in two stages: a) a general description of the specific features of the representation we wish to explain, and the way they arise; b) an attempt to define these features in a formally rigorous way, and to deduce them from postulated features of the metarepresentation. Finally I will try to sketch the features that can as yet not be deduced in this way.

12.2. Basic Concepts and Assumptions of the Metarepresentational Framework

a) The phenomena we wish to represent within this framework are *representations* and *representation processes*. Representations are (actual or potential) abstract structures which allow us to classify and to anticipate external phenomena. Representation processes are (simple or complex) information

transfers happening within a representation or connecting this representation to other representations.

The basic elements of representation structures are postulated to be distinctions. A distinction can be intuitively understood as something allowing to separate a class a of phenomena from its complement or negation a'(i.e. all the phenomena which do not belong to a). According to the particular side (a or a') of the distinction which is actualized the representation will be in a different state. The global state of a representation can then be defined as the conjunction of all the actualized sides of the distinctions constituting the representation.

A representation process can then be understood as something sending present states onto future states of the same or of a different representation. In order to determine which states will be actualized from a given state you should know how the distinctions defining the initial state are related to the distinctions defining the final state. This relation may be called a "causal relation" and can be represented as a morphism of representations, i.e. a mapping preserving a minimal part of the structure (defined by the distinctions) of the representation.

b) In order to formalize this framework we should first express the axioms determining the static properties of distinctions. A very simple axiom system was proposed by Spencer-Brown (1969). His "distinction algebra" is based on just one operator (distinction) (implicitly) defined by two axioms (see sect. 11.2).

It can be shown that this formalism is equivalent to the much more complex formalism of Boolean algebra which is based on four operators (negation, conjunction, disjunction, and implication), and two constants (O and I), and for which several axiom systems exists. Since the latter formalism is better known, I will use these four operators (symbolized respectively by the signs "' ", ". ", "OR ", and "< ", (cf. sect. 4.4 and 4.5), and assume that they obey the usual axioms and theorems of Boolean algebra.

The static part of a representation will then be formally represented by a set of variables (a, b, c, …) together with all their Boolean combinations (e.g. a.b, b OR c', a.(b < c) …). I will further assume that the Boolean algebra defined in this way is atomic. The atoms correspond to the maximal non-zero conjunctions of variables or their negations. Therefore, they can be interpreted as representing the states of the representation.

The distinction algebra of Spencer-Brown and its Boolean equivalent are purely static: they cannot represent change. In order to model representation processes, a further structure must be added: an algebra of morphisms. One way to express this axiomatically is the following:

We can presuppose the axioms of set theory and view a distinction algebra as a set of expressions provided with certain operations. A morphism can then be defined as a function f from an algebra A to an algebra B which preserves the basic operations :

f: A → B: a → f(a) such that: $f(a') = (f(a))'$
$f(a.b) = f(a).f(b)$

(the preservation of the other operators and constants can be deduced from these two requirements).

Theorem: the set of all Boolean algebras together with the set of all Boolean morphisms provided with the operation of composition forms a category (cf. sect. 11.4).

Proof:
- the composition of morphisms considered as functions is associative.
- the function which sends every element of a Boolean algebra B onto itself is a morphism id_A which possesses the identity property: for every f: A → B and for every g: C → A we have:

$id_A * g = g$ and $f * id_A = f$ ■

Note: we could also *postulate* the existence of a category whose objects are Boolean algebras and whose morphisms are defined by the axioms of category theory. In this way no set-theoretical axioms are to be assumed, and hence the axiomatic system remains simpler. In this case, however, it remains to be proven that no additional axioms must be added in order to capture the fundamental properties of the mapping of distinctions we wish to formalize.

On the other hand, such an axiomatization would be closer to the general philosophy underlying the present attempt at formalization (cf. sect. 11.1 and Heylighen, 1984), which aims at a maximal domain of application resting on a minimal set of presuppositions. A consequence of this philosophy is that the proofs of theorems are normally very simple, often so simple that they appear trivial. This should not, however, be considered as pointing to a lack of mathematical sophistication, but rather as an indication that the concepts and postulates were initially well-chosen.

Interpretation*:* the category of all Boolean objects and Boolean morphisms can be viewed as representing the general metarepresentational framework; specific subcategories (i.e. subsets of objects and morphisms which are closed under composition and which contain all the required identities) can be interpreted as particular representation structures (e.g. the classical frame). Functors which map one subcategory onto another one can be viewed as transformations of these structures.

In order to make this categorical algebra into an adaptive metarepresentation we should add one (or more) axioms determining "dynamical constraints" on the representation processes within the framework. For the moment we have not got a sufficient insight into the dynamics of distinctions to do this in an exhaustive way. However, there is one constraint which appears to be necessary in any

case: consistency. In the static Boolean framework this assumption is already implicitly present:

no two different states of the same Boolean object can be jointly actualized: for every $s_1 \neq s_2$ (atoms): $s_1.s_2 = O$.

This could be extended to morphisms by introducing (provisionally) the following requirement:

Consistency axiom: if two morphisms f: A → B and g: A → B are inconsistent, i.e. there is an element a in A such that: f(a).g(a) = O, then f and g cannot both be actual.

Interpretation: this requirement may be viewed as a "conservation of consistency" by morphisms: if two propositions b and c were initially consistent, i.e. there is a proposition a = b.c ≠ O, then they remain consistent after the actual morphisms were applied: f(b).g(c) ≠ O.

12.3. The Classical Representation Frame.

We will here summarize the analysis of chapters 4 to 6 by characterizing the structure of "classical" representations by means of three basic conditions. This will simplify the analysis of the basic non-classical representation structures, since we will show that each non-classical representation is characterized by the violation of one of those three requirements. This does not mean that these conditions are assumed to be sufficient to define a "classical" representation. We will here only show that they are necessarily entailed by the more general condition of "absolute distinction invariance" (ADI), which is assumed to be the necessary and sufficient condition for "classicality".

Each adaptive representation can be analyzed from three basic viewpoints:
1) the logical structure of the state space (or more generally of the set of expressions);
2) the structure of time (and its relation to state space).
3) the algebraic structure of the set of dynamical operators.

In the classical frame these structures are very simple and regular:
1) the relation between states is trivial: either two states are identical or they are orthogonal, i.e. the one implies the negation of the other one.
2) the relation between events (i.e. the states which are actualized) is a complete or linear order: either two events are simultaneous (this means that they can be identified as representing different aspects of the same state of the global system) or the one precedes the other one.
3) the possible operators form a group: they can be inverted and composed with each other without restraint.

We must now show that these features (which will be designated by TO (trivial orthogonality), LT (linear time) and GO (group of operators) respectively) are directly implied by the ADI principle. Applied to the metarepresentational formalism of the preceding section this principle can be formulated as follows :

Definition: ADI holds *iff* the only morphisms which are allowed are those which conserve all distinctions, i.e. the complete structure of the Boolean algebra.

This means that all morphisms are isomorphisms, i.e. they are one-to-one into, each distinct element is sent upon a distinct element. We can then define the relation of isomorphy between Boolean objects: two objects are isomorphic if there exists an isomorphism from the one to the other. This is clearly an equivalence relation. Therefore, all objects of an ADI representation belong to the same equivalence class.

This means that in practice these objects can be identified, since their (internal) structures and their behavior with respect to (external) information processes are the same. Therefore, we may assume that a representation characterized by an ADI requirement consists of only one Boolean algebra provided with its automorphisms.

Theorem: if a representation has the ADI property, then it also has the TO property.

Proof*:* if a representation consists of only one object then its states are those of this object. Since an individual object is assumed to have a Boolean structure, its states are Boolean atoms, and these are necessarily orthogonal. ■

Theorem: if a representation has the ADI property, then it also has the LT property.

Proof: whereas the structure of state space is determined by the possibility of two states to be jointly actual (which possibility is excluded by TO except in the trivial case of the states being identical), the structure of time is determined by the possibility of two representation processes (corresponding to two morphisms) to be jointly actual. In order to check this we should apply the consistency axiom to the automorphisms characterizing an ADI representation.

Suppose we have two different automorphisms f, g: A → A.

It is a basic property of atomic Boolean algebras that every Boolean element can be written as a disjunction of atoms. Since disjunctions are by definition conserved by morphisms, this means that we can restrict our attention to the action of f and g on the set of atoms. f≠ g implies that there exists an atom s, such that:

$f(s) \neq g(s)$.

Since f and g are morphisms, they send atoms onto atoms. Therefore, f(s) and g(s) are atoms and this implies:

$$f(s).g(s) = O.$$

This is in contradiction with the consistency axiom, unless we assume that f and g cannot be jointly actual. This implies that f and g can only be subsequently actual. Therefore, the different actual morphisms describing the changes of the state of an ADI representation form a linear sequence (i.e. no branching occurs), which can be interpreted as describing the order of time. ■

Theorem*:* if a representation has the ADI property, then it also has the GO property.

Proof: the set of all automorphisms of a Boolean object forms a group under composition. Indeed, since there is only one object which is both domain and co-domain of the morphisms, composition is everywhere defined. The identity morphism is an automorphism, and the inverse of an automorphism is again an automorphism (the ADI property is indeed a conservation principle, hence it is invariant for inversion). ■

Conclusion: we have proven the following propositions:
1) ADI < TO
2) ADI < LT
3) ADI < GO
Therefore, we have also proven the contraposition of these implications:
1') NOT (TO) < NOT (ADI)
2') NOT (LT) < NOT (ADI)
3') NOT (GO) < NOT (ADI).

The first three propositions can be used to characterize the classical frame, i.e. the representation structure determined by ADI. We will now examine the three non-classical representation structures, i.e. structures characterized by the respective negation of TO, LT and GO, and hence by the negation of ADI.

12.4. The Quantum Mechanical Frame.

As was shown in sect. 7.2 the representation structure of quantum mechanics can be seen as an illustration of proposition (1"): the logical structure of the quantum state space (the "Hilbert space") corresponds to the specific feature of the quantum observation process (described as the "collapse of the wave function", and formalized by the "projection postulate"), which is such that different states can be sent upon the same state, contradicting the ADI principle. Let us analyze this non-ADI feature in more detail.

This non-classical nature of quantum representations can be characterized in many different ways (chapter 7): complementarity of subrepresentations, indeterminacy, superposition of states, non-Boolean logic, non-Bayesian probability, … We have summarized these descriptions by stressing the non-TO structure of quantum mechanics, which is expressed most sharply by the "logical" version of the superposition principle:

Superposition principle: for every two different states s_1, s_2 there exists a third state s_3 which is orthogonal neither to s_1 nor to s_2 (i), but which is orthogonal to all states orthogonal to both s_1 and s_2 (ii).

We must now examine in how far this principle can be deduced from the metarepresentational framework and from the basic properties of the observation process (viewed as an information transfer between a microscopic phenomenon and a macroscopic observation apparatus). Let us begin with the observation process.

We will consider the observer together with his observation apparatus (which can be viewed as an extension of his natural sense organs) as a basic adaptive or cognitive system. This means that the information processed by the observer is necessarily filtered, hence incomplete. This was expressed by the principle that "every distinction implies an assimilation". This principle can be viewed as the consequence of an even more basic principle stating that a representation cannot completely represent its own interaction with the outside world, more in particular that it cannot have a complete "self-knowledge".

The filtering of information itself is a relatively simple phenomenon which is not restricted to quantum observations. For example, in statistical mechanics observations result in "macrostates" which are by definition incomplete representations of phenomena determined at the microscopic level.

What is typical of quantum processes is that the interaction effect which goes back from the apparatus to the micro-object which was observed, cannot be neglected. Since the observer cannot have a complete knowledge of this effect, the perturbation of the state of the object will be partly indeterminate. Therefore, each observation creates a specific indeterminacy, or, in the present terminology, wipes out a specific set of potential distinctions characterizing the phenomenon.

As shown by the Bohr thought experiment, the particular distinctions which are wiped out depend upon the observation set-up (sect. 7.3). You could say that each observation set-up is characterized by a particular (incomplete) representation. On the other hand different set-ups, hence different incomplete representations, are generally incompatible: they cannot be coordinated in one integrated set-up, but they cannot be applied subsequently either, because the first one would create an indeterminacy perturbing the results of the next one. Therefore, we are left with a set of incomplete but mutually incompatible ("complementary") representations, each describing the potential observation results (distinctions) of a collection of (compatible) set-ups.

Let us now try to reformulate these rather intuitive ideas by using the metarepresentational formalism. We shall assume that the (incompletely

observable) microstate of the object could be represented by a classical "microscopic" representation, i.e. by a single Boolean algebra C. The distinctions belonging to this representation can be interpreted as representing all the observable and hidden variables determining the quantum system.

The observation process can then be represented by an epimorphism (cf. sect. 11.4) f, sending this complete microscopic representation onto an incomplete macroscopic representation. The macroscopic representation is again a Boolean algebra M, whose distinctions represent the macroscopically distinguishable states of a collection of compatible observation set-ups.

For example, in the most simple case M would consist of only one distinction: that between spin-up and spin-down, which is registered as the difference between the detection of a particle behind a polarizer, and the absence of that detection (signifying that the particle has not passed the polarizer). In that case we have:

$f: C \rightarrow M: a \rightarrow f(a)$, where f(a) is either u (spin-up) or u' (spin-down).

We may then define $f^{-1}(u)$ as the set of all microvariables (or simpler microstates) in C which are sent by f onto the macrostate u. Since the information transferred through f is incomplete, this set contains more than one element (f is many-to-one). Therefore, f^{-1} is not a regular function.

However if we now wish to represent the possible macro-distinctions made by another incompatible set-up (e.g. one for measuring the spin in the left-right direction), then we must consider another macrorepresentation and another observation epimorphism:

$f_1: C \rightarrow M_1$ e.g. where: $M_1 = \{l, l'\}$

Now f_1^{-1} (l) corresponds to a different subset of C (or of its state space). Since we can assume that both f and f_1 will each send half of the states of C onto u, respectively onto l, and the other half onto u', respectively l', the subsets corresponding to these functions will not be disjoint. Their intersection corresponds to the set of microstates which would be sent to u by f and to l by f_1.

However, since the representations are incompatible this intersection cannot be observed, it corresponds to a proposition of the unobservable micro-representation C, it does not correspond to an observable distinction. We will therefore assume that the conjunction of the two observable propositions is zero, even though the intersection of the corresponding subsets of microstates is not zero.

Definition: $S_m = \{s \in S: f \text{ observation epimorphism: } f(s) = m\}$

Interpretation: S_m can be interpreted as the minimal set of (orthogonal) microstates which can be sufficiently enlarged by a single collection of

compatible observation set-ups to become macroscopically distinguishable as a proposition m of the macroscopic subrepresentation M. If m is an atom of M, then the elements of S can be said to represent the hidden variables of the observed system. These hidden variables, however, are not subjected to the same restrictions as the traditional hidden variables since the macrostates m viewed as subsets of hidden variables are not disjoint.

This allows us to define a non-trivial orthogonality relation between macrostates (or macro-propositions):

Definition: $a \perp b$ *iff* $S_a \cap S_b = \varnothing$

According to this definition two macrostates will be orthogonal if they belong to the same classical subrepresentation determined by the same observation epimorphism f. Macrostates belonging to different classical subrepresentations will not be orthogonal since different distinctions are filtered out.

This construction also allows us to define a non-classical transition probability between macrostates:

Definition: $$P(m_1 \mid m_2) = \frac{N(S_{m1} \cap S_{m2})}{N(S_{m2})}$$

where N is a measure on the microstate space.

Property: P does not obey the Bayes axiom for conditional probability. Indeed, for $S_{m1} \cap S_{m2} \neq \varnothing$:

$$P(m_1 \mid m_2) \neq P(m_1.m_2) / P(m_2) = P(O) / P(m_2) = O$$

Theorem: the set of all macrostates (defined in this way) which encompasses the atoms of all maximally allowed classical subrepresentations, satisfies the superposition principle.

Proof: consider two different macrostates m_1 and m_2, characterized by the sets S_{m1} and S_{m2}, which are their inverse images through observation morphisms.

We can then always find a third state m_3 which is determined by a minimally distinguishable subset $S_m \subset S_{m1} \cup S_{m2}$.

Then m_3 is orthogonal neither to m_1 nor to m_2, since

$S_m \cap S_{m1} \neq \varnothing$ and $S_m \cap S_{m2} \neq \varnothing$

Moreover m_3 is orthogonal to all states m_4 orthogonal to m_1 and m_2: suppose $m_4 \perp m_1$ and $m_4 \perp m_2$, then:

$S_{m4} \cap (S_{m1} \cup S_{m2}) = \varnothing$, but then also: $S_{m4} \cap S_m = \varnothing$ and hence $m_4 \perp m_3$. ■

Note: this construction clearly captures the basic non-classical feature of quantum mechanics: the superposition principle (in a quantum logical context this principle is sometimes formulated as "the atomic bisection property" or "the irreducibility" of a quantum lattice (Piron, 1976; Ivert and Sjödin, 1978)). It is easily extended so that it further captures the basic properties which quantum logic inherits from Boolean logic: completeness of the lattice, atomicity, orthocomplementation and orthomodularity (see Ivert and Sjödin, 1978 for a complete list of axioms).

However, there remains one axiom determining quantum logics which for the moment does not seem to follow from the construction: the (atomic) covering law. In order to explain this feature we should further specify adequate restrictions on the non-atomic components of the lattice, e.g. by means of a "closure" operation (cf. sect. 11.3).On the other hand it has been shown by Aerts (1981) that the covering law cannot be satisfied by a system consisting of two non-classical separated entities. It thus remains to be studied in how far the covering law can or should be introduced in the present construction in order to model all essential properties of quantum representations.

12.5. The Relativistic Frame.

Relativity theory is characterized by the negation of the LT condition, hence by the violation of the ADI principle: temporal precedence is no longer a linear order relation but a partial order. This is analogous to quantum mechanics: a strong property (TO or LT) is negated and replaced by an at first sight weaker property (not all states are orthogonal or not all events can be temporally ordered). However, if we look more closely at the fundamental structures of these two theories, we find an additional property which is missing in the classical frame.

In quantum mechanics, this can be found in the superposition principle which does not only express that there are non-orthogonal states (i), but also that these states can somehow be grouped in "subspaces" such that they are together orthogonal to the same other "subspaces" of states (ii). It is this requirement which provides the quantum state space with an additional "coherence" which is lacking in classical state spaces.

An analogous phenomenon can be found in relativistic "space-time": space-time is more than just an incompletely ordered set of events, it possesses a so-called "causal structure", which is determined by *two* distinct relations, chronological precedence and horismotic precedence. This defines a topology on space-time (in the classical frame, on the other hand, the topology must be brought in from the outside, without clear motivation).

In the present approach we attempt to explain these additional structures by looking at the non-ADI morphisms connecting different representations which are ignored by the classical frame. Indeed, if we do not require ADI, then the

representation structure will consist of different, non-isomorphic Boolean objects, connected by generally non-invertible morphisms. Each object can be interpreted as representing the potential states of a particular system at a particular instant in time. The actualization of a particular state of an object can be called an "event".

(In the classical frame we did not have to introduce events because all objects were isomorphic and hence could be identified. In that way it suffices to introduce one global state representing all the different subsystems. Events are then merely subsequent global states.)

We should now examine the structure of the set of events. This structure is determined by the graph of all the morphisms which can be jointly actualized. This actualization is constrained by the consistency axiom. For the moment, we will not study the detailed dynamical structure entailed by this axiom, we will only make a rough classification of the different paths formed by the concatenation of morphisms.

Definition: a set of morphisms f: A → B, g: B → C, h : C → D, …. will be called a "path" if they can be sequentially composed, i.e. there exists a morphism: p = … ∗ h ∗ g ∗ f.

Such a path can be interpreted as a "(globally) causal relation" (in the sense of sect. 8.5). Indeed the composability of the (local) morphisms f, g, h, … signifies that they allow a (global) morphism p, which still transfers at least one distinction from the initial object A to the final object (otherwise it would not be a morphism). The consistency axiom can now be applied to compare "parallel" paths, i.e. paths whose initial and final objects coincide (although the intermediate objects may be different). Let us apply it first to cyclic paths.

Definition: a path p: A → Z is cyclic if its initial and final objects coincide: A = Z.

Property: a cyclic path has at least one parallel path: the identity morphism of its initial object: id_A.

Theorem: either a cyclic path p: A → A is inconsistent with the identity morphism of A and then it can never be actual, or it sends states onto weaker propositions: s ∈ A atom, then s < p(s).

Proof*:* suppose the state s is actual. Then we have two possible cases:
1) p(s).s = O
2) p(s).s≠ O.

1) In the first case we have: $p(s).id_A(s) = p(s).s = O$.

Therefore, p is inconsistent with the identity morphism, which signifies according to the consistency axiom that p and the identity can not be jointly actual. But since by definition every categorical object must have a

corresponding identity morphism we cannot exclude id_A and hence we must exclude p.
2) Since s is a Boolean atom, this case implies that: p(s).s = s, and hence: s < p(s). ■

Interpretation: cyclic paths clearly cannot be used to transfer information: either they cannot be actual or the information they convey (that p(s) should be actual) is already contained in the state. This means that if we wish to represent potentially information transferring processes then we should restrict ourselves to non-cyclic paths. These paths can be further divided in two distinct classes:

Definition: an acyclic path is "chronological" if it has at least one different, parallel path (which is evidently not an identity morphism).

Definition: a path is "horismotic" if it has no parallel path.

Remark: clearly a horismotic morphism can never be inconsistent, simply because there is no other path to be inconsistent with; for chronological paths however the different conditions for consistency will have to be further analyzed.

Theorem: the set of all (relativistic) objects together with the relations CP, HP and P, which are determined by respectively the chronological paths, the horismotic paths and the union of both sets of paths, form a "causal space" (as defined by Kronheimer and Penrose, 1967).

Proof: the proof is analogous to that of sect. 8.5, with "events" replaced by "objects" and "causal connections" by "morphisms". ■

Note: the causal structure can be viewed as the most fundamental feature of relativistic space-time. However, it is not the only one. If we wish to completely reconstruct space-time then we should also be able to determine features such as the metric or the dimension of space-time, which in the existing theories are introduced axiomatically. A possible approach to "derive" these properties from the properties of information transmissions between events is hinted at in (Heylighen, 1984) and in sect. 8.3, page 109 (the Marzke-Wheeler construction). It is clear, however, that a lot of research is still needed to elucidate these problems.

12.6. The Thermodynamic Frame.

The non-classical feature of the set of theories we roughly denote by the term "thermodynamics", resides in the irreversibility of the processes described within their representation. Therefore, the operators used to represent such processes cannot form a group (in the best case only a monoid or semigroup).

Therefore, the representation is characterized by the negation of the GO property and hence, once again, by the negation of ADI. In order to represent this in the metarepresentational framework, it suffices to consider not only automorphisms (representing the reversible transformations) but also endomorphisms (representing internal processes with a loss of distinctions, i.e. an increase of entropy).

However, it would be interesting to find which "additional structure" characterizes thermodynamic evolution. Since the thermodynamic theories are much less "well-formalized" than quantum mechanics or relativity theory, it is at the moment not clear which structures are really fundamental. We can only hope that the metarepresentational framework will prove to be a useful guide in the further analysis of irreversible processes.

12.7. Conclusion

We have described a possible way to formalize the metarepresentational framework, whose basic ideas were explained on a more intuitive level in the thesis. The formalism is based on a fairly simple collection of postulates: the axioms of distinction algebra (or equivalently of Boolean algebra), the axioms of set theory (possibly replaced by those of category theory), and finally a newly introduced axiom requiring the consistency of parallel morphisms.

With this axiom system we have shown that the basic properties of the classical representation frame could be derived from the requirement of absolute distinction invariance: orthogonality of states, linearity of time and the group structure of the operator algebra.

The three basic, non-classical representation structures (quantum mechanics, relativity theory and thermodynamics) could then be characterized by the respective negation of these three properties. Finally, we have shown that in the case of quantum mechanics and relativity theory additional structures could be deduced: respectively the structure entailed by the superposition principle, and the causal structure of space-time.

BIBLIOGRAPHY

Aerts, D. (1982): "Description of Many Separated Physical Entities without the Paradoxes Encountered in Quantum Mechanics", Found. Phys. 12, p. 1131.

Aerts D. (1982): "Example of a Macroscopic System that Violates Bell Inequalities", *Lett. Nuovo Cimento 34*, p.107.

*Aerts D. (1983): "Classical Theories and Non Classical Theories as Special Cases of a More General Theory", *J. Math. Phys. 24*, p.2441.

*Aerts D. (1986): "A Possible Explanation for the Probabilities of Quantum Mechanics", *J. Math. Phys. 27*, p.202.

Aharonov Y. and Bohm D. (1959): "Significance of Electromagnetic Potentials in the Quantum Theory", *Phys. Rev. 115*, p.485.

*Amarel, S. (1968): "On Representations of Problems of Reasoning about Actions", in: *Machine Intelligence 3*, D. Michie (ed.), (American Elsevier, New York), p.131.

Apostel L. (1975): "Matière et Forme: introduction à une épistémologie réaliste", (Communication & Cognition, Ghent).

Arbib M.A. and Manes E.G. (1975): "Arrows, Structures and Functors", (Academic Press, New York).

Aspect A., Dalibard J. and Roger G. (1982): "Experimental Test of Bell's Inequalities Using Time-Varying Analyzers", *Phys. Rev. Lett. 25*, p.1804.

Beeson M. J. (1985): "Foundations of Constructive Mathematics" (Springer, Berlin)

Bell J.S. (1966): *Rev. Mod. Phys. 38*, p.827.

*von Bertalanffy L. (1968): "General Systems Theory", (Braziller, New York).

Birkhoff G. and Von Neumann J. (1935): *Annals of Mathematics 37*, p.823.

Blatt S.J. (1984): "Continuity and Change in Art: the Development of Modes of Representation", (Erlbaum, Hillsdale).

Bobrow D.G. (1975): "Dimensions of Representation" in: Bobrow and Collins (1975), p.1.

Bobrow D.G. and Collins A. (eds.) (1975): "Representation and Understanding. Studies in Cognitive Science ", (Academic Press, New York).

Bohr N. (1958): "Atomic Physics and Human Knowledge", (Wiley, New York).

*de Bono E. (1969): "The Mechanism of Mind", (Penguin, Harmondsworth).

Briggs J.P. and Peat F.D. (1984): "Looking Glass Universe. The Emerging Science of Wholeness", (Simon and Schuster).

Brillouin L. (1962): "Science and Information Theory", (2nd. ed., Academic Press, New York).

de Broglie L. (1959): *Journ. Phys. Rad. 20*, p.963.

Burghgraeve P. (1976): "On Representations of Problems in Heuristic Problem-Solving", *Communication & Cognition 9*, p.231.

*Campbell D.T. (1974): "Evolutionary Epistemology", in: "*The Philosophy of Karl Popper*", Schilpp P.A. (ed.), (Open Court Publishing, La Salle, Ill.), p.413.

Carnap R. (1958): "Introduction to Symbolic Logic and its Applications", (Dover, New York).

*Charniak E. and McDermott D. (1985): "Introduction to Artificial Intelligence", (Addison-Wesley, Reading, Mass.).

Chomsky N. (1972): "Language and Mind", (Harcourt Brace Jovanovich, New York).

Cohen H. (1979): "What is an Image?" in: *Proceedings of the 6th International Joint Conference on Artificial Intelligence*.

*Cohen H. (1982): "On the Modelling of Creative Behavior", *Rank Xerox Corp.* (internal report).

Despontin M. (1986): "Modèles d'aide à la décision" in: *Proceedings of the 11th International Congress on Cybernetics* (Namur).

Einstein A., Podolsky B. and Rosen N. (1935): *Phys. Rev. 47*, p.1804.

Farmer D., Toffoli T. and Wolfram S. (eds.) (1984): "Cellular Automata: Proceedings of an Interdisciplinary Workshop", (North Holland Physics Publishing, New York).

Ferguson M.(1982): "The Aquarius Conspiracy" (Granada Publishing, London).

*Finkelstein D.: "Space-Time Code I; II; III; IV; V ", *Phys. Rev. 184* (1969), p.1261; *D5* (1972), p.320; *D5* (1972), p.2922; *D9* (1974), p.2219; *D9* (1974), p.2231.

*Finkelstein D. (1979): "Holistic Methods in Quantum Logic", in: *Quantum Theory and the Structures of Space and Time 3*, Castell L.,Drieschner M. and von Weizsacker C.F. (eds.), (Hanser, Munich).

Forbus K.D. (1981): "Qualitative Reasoning about Physical Processes", in: *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, p.326.

Galbraith J.R. (1977): "Organization Design", (Addison-Wesley, Reading, Mass.).

Ghirardi G.C., Rimini A. and Weber T. (1980): "A General Argument against Superluminal Transmission through the Quantum Mechanical Measurement Process", *Lett. Nuovo Cimento 27*, p.293.

Gille J.C. and Clique M. (1975): "La représentation d'état pour l'étude des systèmes dynamiques", (Ed. Eyrolles, Paris).

Haken H. (1978): "Synergetics: Non-Equilibrium Phase Transitions and Self Organization in Physics, Chemistry and Biology", (Springer, Berlin).

Halmos P.R. (1974): "Lectures on Boolean Algebras", (Springer, New York).

*Hawkes T. (1977): "Structuralism and Semiotics", (Methuen, London).

*Hayes P. (1978): "The Naive Physics Manifesto" in: Michie D. (ed.): *Expert Systems in the Microelectronic Age*, (University of Edinburgh Press).

Heeffer A. (1986): "Multiple Strategies for Knowledge Manipulation Using Meta-Level Reasoning" in: *Proceedings of the 11th International Congress of Cybernetics*, (Namur).

*Heylighen F. (1984): "A Structural Language for the Foundations of Physics", preprint VUB/TF/84/10 (Brussels).

Heylighen F. (1986): "Towards a General Framework for Modelling Representation Changes" in: *Proceedings of the 11th International Congress on Cybernetics*, (Namur).

Heylighen F. (1987): "Formal Foundations for an Adaptive Metarepresentation", to be published in: *Proceedings of the 7th International Congress of Cybernetics and Systems*, (Thales Publications, St. Annes-on-Sea, Lancs., England).

Ivert P.A. and Sjödin T. (1978): "On the impossibility of a finite propositional lattice for quantum mechanics", Helvetica Physica Acta 51, p. 635.

*Jammer M. (1974): "The Philosophy of Quantum Mechanics", (Wiley, London).

Van Kampen N. (1985): "Can the Aharonov-Bohm Effect Transmit Signals Faster than Light?", *Phys. Lett. 106A*, p.5.

Kauffman S.A. (1984): "Emergent Properties in Random Complex Automata", *Physica D 10*, p.145.

de Kleer J. and Brown J.S. (1983): "A Qualitative Physics based on Confluences", in: Hobbes J. and Moore B. (eds.): *Formal Theories of the Common-Sense World*, (Ablex).

*Korf R.E. (1980): "Toward a Model of Representation Changes", *Artificial Intelligence 14*, p.41.

*Kronheimer E.H. and Penrose R. (1967): "On the Structure of Causal Spaces", *Proc. Camb. Phil. Soc. 63*, p.481.

Kuhn T.S. (1962): "The Structure of Scientific Revolutions", (University of Chicago Press, Chicago).

Lenat D.B. (1983): "EURISKO: A Program that Learns New Heuristics and Concepts", *Artificial Intelligence 21*, p.61.

MacLane S. and Birkhoff G. (1967): "Algebra", (Collier-MacMillan, London).

March J.G. and Simon H.A. (1958): "Organizations", (Wiley, London).

Maturana H.R. and Varela F.J. (1980): "Autopoiesis and Cognition: the realization of the living", (Reidel, Dordrecht).

Mesarovic M.D., Macko D. and Takahara Y. (1970): "Theory of Hierarchical, Multilevel Systems", (Acad. Press, New York).

*Mesarovic M.D. and Takahara Y. (1975): "Mathematical Foundations of General Systems Theory", (Academic Press, New York).

*De Mey M. (1982): "The Cognitive Paradigm" (Reidel, Dordrecht)

*Misner C.W., Thorne K.S. and Wheeler J.A. (1974): "Gravitation", (Freeman, San Francisco).

Misra B. and Sudarshan E.C.G. (1977): "The Zeno's Paradox in Quantum Mechanics", *J. Math. Phys. 18*, p.756.

*Nauta D. (1972): "The Meaning of Information", (Mouton, Den Haag).

Newell A. and Simon H.A. (1972): "Human Problem Solving", (Prentice Hall, Englewood Cliffs).

Nilsson N.J. (1971): "Problem-Solving Methods in Artificial Intelligence", (McGraw-Hill, New York).

Piaget J. (1972): "L'épistémologie génétique", (PUF, Paris).

Piron C. (1976): "Foundations of Quantum Physics", (W.A. Benjamin, Menlo Park, Calif.).

Podlaha M.F. (1980): "On the Impossibility of Measuring the One-Way Velocity of Light", *Lett. Nuovo Cimento 28*, p.216.

Popper K. (1968): "The Logic of Scientific Discovery", (Rev. Ed., Hutchinson, London)

*Prigogine I. (1979): "From Being to Becoming: Time and Complexity in the Natural Sciences", (W.H. Freeman, San Francisco).

Prigogine I. and Stengers I. (1979): "La Nouvelle Alliance: métamorphose de la science", (Gallimard, Paris).

*Reichenbach H. (1958): "The Philosophy of Space and Time", (Dover, New York).

Van Releghem E. (1985): "Reformulation of Knowledge", (AI lab, VUB, Brussels, internal report).

Rindler W. (1977): "Essential Relativity", (2nd. ed., Springer, New York).

Roy B. (1985): "Méthodologie multicritère d'aide à la décision", (Economica, Paris).

Saletan E.J. and Cromer A.H. (1971): "Theoretical Mechanics", (Wiley, New York).

Sikorsky R. (1969): "Boolean Algebras", (3rd ed., Springer, New York).

Simon H.A. (1957): "Models of Man: Social and Rational", (Wiley, New York).

*Simon H.A. (1977): "Models of Discovery", (Reidel, Dordrecht).

*Simon H.A. (1979): "The Sciences of the Artificial", (2nd Ed., MIT Press, Cambridge, Mass.).

Sjödin T. (1979): "Synchronization in Special Relativity and Related Theories", *Nuovo Cimento 51 B*, p.229.

Sjödin T.and Heylighen F.(1985): "Tachyons Imply the Existence of a Privileged Frame", *Lett. Nuovo Cimento 44*, p.617.

*Spencer Brown G. (1969): "Laws of Form" (Allen and Unwin, London)

Steels L. (1984): "Object Oriented Knowledge Representation in KRS" in: *Proceedings of the 6th European Conference on Artificial Intelligence*, O'Shea T. (ed.), (Elsevier Science Publishers, Amsterdam).

Toffler A. (1970): “Future Shock” (Random House, New York).

*Varela F.J. (1979): “Principles of Biological Autonomy”, (North Holland, New York).

Watzlawick P., Weakland J.H. and Fisch R. (1975): “Changes: Paradoxes and Psychotherapy”.

Wickelgren W.A. (1977): “Learning and Memory”, (Prentice-Hall, Englewood Cliffs, N.J.).

Wickelgren W.A. (1974): “How to Solve Problems: Elements of a Theory of Problems and Problem-Solving”, (Freeman, San Francisco).

Wheeler J.A. (1968): “Superspace and Quantum Geometrodynamics” in: De Witt C.M. and Wheeler J.A. (eds.): *Batelle Rencontres 1967: Lectures in Mathematics and Physics*, (Benjamin, New York).

Wolfram S. (1984): “Universality and Complexity in Cellular Automata”, *Physica 10 D*, p.1.

*Woodhouse N. M. J. (1973): “The Differentiable and Causal Structures of Space-Time”, *J. Mat. Phys. 14*, p. 495.

Zeeman E.C. (1964): “Causality Implies the Lorentz Group”, *J. Math. Phys. 5*, p.490.

(the * sign indicates references which were particularly inspiring during the research which led to the present volume)

Papers in which some ideas of the book have been further elaborated

Heylighen F. (1989): “Causality as Distinction Conservation: a theory of predictability, reversibility and time order”, *Cybernetics and Systems* 20, p. 361-384.

Heylighen F. (1990): “Non-Rational Cognitive Processes as Changes of Distinctions”, *Communication & Cognition* 23, No. 2-3, p. 165-181.

Heylighen F. (1990): “Classical and Non-classical Representations in Physics I”, *Cybernetics and Systems* 21, p. 423-444.

Heylighen F. (1990): “Classical and Non-classical Representations in Physics II: Quantum mechanics”, *Cybernetics and Systems* 21, p. 477-502.

Heylighen F. (1990): “A Structural Language for the Foundations of Physics”, *International Journal of General Systems* 18, p. 93-112.

Heylighen F. (1991): “Modelling Emergence”, *World Futures: the Journal of General Evolution* 31 (Special Issue on *Emergence*, edited by G. Kampis), p. 89-104.

Heylighen F. (1991): “Coping with Complexity. Concepts and principles for a support system”, *Systemica* 8, part 1 (special issue on *Mutual Uses of Cybernetics and Science*, edited by Glanville R. & de Zeeuw G.), p. 39-55.

Heylighen F. (1992): “From Complementarity to Bootstrapping of Distinctions: A Reply to Löfgren's Comments on my Proposed 'Structural Language”, *International Journal of General Systems* 21, no. 1, p. 99.

Heylighen F. (1986): “Towards a General Framework for Modelling Representation Changes”, in: *Proc. 11th Int. Congress on Cybernetics*, (Assoc. Internat. de Cybernétique, Namur), p. 136-141.

Heylighen F. (1987): “Formal Foundations for an Adaptive Metarepresentation”, in: *Cybernetics and Systems: the Way Ahead V2* J. Rose (ed.), (Thales, St. Annes-on-Sea, Lancashire), p. 648-651.

Heylighen F. (1988): “Formulating the Problem of Problem-Formulation”, in: *Cybernetics and Systems '88*, Trappl R. (ed.), (Kluwer Academic Publishers, Dordrecht), p. 949-957.

Heylighen F. (1990): “Autonomy and Cognition as the Maintenance and Processing of Distinctions”, in: *Self-Steering and Cognition in Complex Systems*, Heylighen F., Rosseel E. & Demeyere F. (ed.), (Gordon and Breach, New York), p. 89-106.

Heylighen F. (1990): “Relational Closure: a mathematical concept for distinction-making and complexity analysis”, in: *Cybernetics and Systems '90*, R. Trappl (ed.), (World Science, Singapore), p. 335-342.

Heylighen F. (1992): “Distinction Dynamics: from mechanical to self-organizing evolution”, in: *Proc. of the Int. Workshop* “*Analysis and Control of Dynamical Systems*”, E. Gindev (ed.), (CLCS, Bulgarian Academy of Sciences, Sofia).

Index